



QY 181 GCGAGACGGGGCGGCTCTGAGACTCCGGGCTCCGCTCTTTCCGGGACCGCCCACTA 240  
Db 181 GCGAGACGGGGCGGCTCTGAGACTCCGGGCTCCGCTCTTTCCGGGACCGCCCACTA 240  
QY 241 CCACGACTCCGACAGAGGGTGAAAAAGATAAATTCGGGTCTCGCGATCGTCTCTAATC 300  
Db 241 CCACGACTCCGACAGAGGGTGAAAAAGATAAATTCGGGTCTCGCGATCGTCTCTAATC 300  
QY 301 TCGCGAAGAGAGAGGGCGGCCCATCGGCCGAACGAGGCGGTGGCGAGGAGGGGGTG 360  
Db 301 TCGCGAAGAGAGAGGGCGGCCCATCGGCCGAACGAGGCGGTGGCGAGGAGGGGGTG 360  
QY 361 TGGCCGGGAGCGCGAAGTCCCCGGGAGTAAGGGAGAGGGGGCGGGTCCGCGTCCCCGG 420  
Db 361 TGGCCGGGAGCGCGAAGTCCCCGGGAGTAAGGGAGAGGGGGCGGGTCCGCGTCCCCGG 420  
QY 421 GCATACGATCGGTGCACGCTGCCGGTCCGGCTGGGCTGAGAGGGGAGGGGGCGGGCG 480  
Db 421 GCATACGATCGGTGCACGCTGCCGGTCCGGCTGGGCTGAGAGGGGAGGGGGCGGGCG 480  
QY 481 GCCGAGCGCGGTCTGTTATTTCCGTGGTCCGGACAGTGGTGGCGGGCGGGGTGACCAAG 540  
Db 481 GCCGAGCGCGGTCTGTTATTTCCGTGGTCCGGACAGTGGTGGCGGGCGGGGTGACCAAG 540  
QY 541 GGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGGATGGAG 600  
Db 541 GGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGGATGGAG 600  
QY 601 CTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATAAACAATGGGTGGACATTTACCAAGAT 660  
Db 601 CTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATAAACAATGGGTGGACATTTACCAAGAT 660  
QY 661 TTTGAAGAGCTTGTGAGAAATTAAGTTGGAGAACTACTTCATGATAAGCTATTGGT 720  
Db 661 TTTGAAGAGCTTGTGAGAAATTAAGTTGGAGAACTACTTCATGATAAGCTATTGGT 720  
QY 721 CTTTGTGAAGCCATGTCTGCTATTGAATGATGGATCCCAAGATGGATGCTGGCATGATT 780  
Db 721 CTTTGTGAAGCCATGTCTGCTATTGAATGATGGATCCCAAGATGGATGCTGGCATGATT 780  
QY 781 GGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT 840  
Db 781 GGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT 840  
QY 841 AAAATTAAGATCTCACCTTGCCCTGAACTGATAGGATTAATGGATACATGTTTTCCTGT 900  
Db 841 AAAATTAAGATCTCACCTTGCCCTGAACTGATAGGATTAATGGATACATGTTTTCCTGT 900  
QY 901 TTGATAACGTTGTTAGAGGCCATTCACTGGCACAGACAGATTTTACGTCCTTTACATT 960  
Db 901 TTGATAACGTTGTTAGAGGCCATTCACTGGCACAGACAGATTTTACGTCCTTTACATT 960  
QY 961 CATAATCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAA 1020  
Db 961 CATAATCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAA 1020  
QY 1021 ATCTGTGACATTCGAAGGGAAAAAGTAAATAAAGCTGCTGTTTGAAGAGGAAGATTTT 1080  
Db 1021 ATCTGTGACATTCGAAGGGAAAAAGTAAATAAAGCTGCTGTTTGAAGAGGAAGATTTT 1080  
QY 1081 CAGTCAATGACTTATGGATTTAAAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGGC 1140  
Db 1081 CAGTCAATGACTTATGGATTTAAAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGGC 1140  
QY 1141 ATGCTAAAAGATGTGGAGGATGACATGCAAAAGAGTAAAGAGTACTCGAAGTCGACAA 1200  
Db 1141 ATGCTAAAAGATGTGGAGGATGACATGCAAAAGAGTAAAGAGTACTCGAAGTCGACAA 1200  
QY 1201 GGAGAAAGAAAGAGATCCAGAAAGTTGAACCTAGAACACCAACAATGTTTAGCAGTATTGAGC 1260  
Db 1201 GGAGAAAGAAAGAGATCCAGAAAGTTGAACCTAGAACACCAACAATGTTTAGCAGTATTGAGC 1260

QY 1261 AGAGTGAAAAATTTACTCGTGTGTTACTGACAGTGTCTATAGCCTTTTACTAAGAAAGAGACC 1320  
Db 1261 AGAGTGAAAAATTTACTCGTGTGTTACTGACAGTGTCTATAGCCTTTTACTAAGAAAGAGACC 1320  
QY 1321 AGTGTGTTGAGAAAGCTCAAAAATTTGATGTTCAAGAGCAGATCTTCTTTCTGCCATT 1380  
Db 1321 AGTGTGTTGAGAAAGCTCAAAAATTTGATGTTCAAGAGCAGATCTTCTTTCTGCCATT 1380  
QY 1381 CATAATTCATTCATCATGCGATCCAGGCCAGAAATGATATCTACAAAAGGAGATCATCCA 1440  
Db 1381 CATAATTCATTCATCATGCGATCCAGGCCAGAAATGATATCTACAAAAGGAGATCATCCA 1440  
QY 1441 ATATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACTACCTTCCCTCGA 1500  
Db 1441 ATATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACTACCTTCCCTCGA 1500  
QY 1501 TATGCAAAAAATAATTAAAGGGAAGAAATGGTGAACCTATTTTGAAGATTAATAGATAGA 1560  
Db 1501 TATGCAAAAAATAATTAAAGGGAAGAAATGGTGAACCTATTTTGAAGATTAATAGATAGA 1560  
QY 1561 ATAAAAACTGTCTGTGAGGTTGTGAATTTAAACAAATTTACATTTCTGTGATTTTTC 1620  
Db 1561 ATAAAAACTGTCTGTGAGGTTGTGAATTTAAACAAATTTACATTTCTGTGATTTTTC 1620  
QY 1621 TGTGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTACAAACCACTTTC 1680  
Db 1621 TGTGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTACAAACCACTTTC 1680  
QY 1681 CTGCTGGATAACAAAAAGGTCCTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCA 1740  
Db 1681 CTGCTGGATAACAAAAAGGTCCTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCA 1740  
QY 1741 CTTCCGCTCTTTTGTGAGATCCTCCGAGTGTCTTCCCAAGTCTACCTATATAATAATC 1800  
Db 1741 CTTCCGCTCTTTTGTGAGATCCTCCGAGTGTCTTCCCAAGTCTACCTATATAATAATC 1800  
QY 1801 ACCAGGCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTCGGCCATTTCTGTAGTC 1860  
Db 1801 ACCAGGCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTCGGCCATTTCTGTAGTC 1860  
QY 1861 TTATTGAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAATTTCTTG 1920  
Db 1861 TTATTGAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAATTTCTTG 1920  
QY 1921 AGGAATTTGCCACCTTGCAGGATGATTTATGACATTTTATTTAATAGGCAGAGAAAGT 1980  
Db 1921 AGGAATTTGCCACCTTGCAGGATGATTTATGACATTTTATTTAATAGGCAGAGAAAGT 1980  
QY 1981 TGATGACGGCTTCACACCATTGCTGTGAACACAGGAACCCCAAGGCAACATTTTGGCCTG 2040  
Db 1981 TGATGACGGCTTCACACCATTGCTGTGAACACAGGAACCCCAAGGCAACATTTTGGCCTG 2040  
QY 2041 TTTAGGTACCTGGGTCTTTACCATAACTTTCGATTTATGATACAGTACCTTCTAAGTGG 2100  
Db 2041 TTTAGGTACCTGGGTCTTTACCATAACTTTCGATTTATGATACAGTACCTTCTAAGTGG 2100  
QY 2101 CTTTGAATTTGAACTCTACAGTATGACAGTACTATTTACATATTTGGTATCTCTCTGA 2160  
Db 2101 CTTTGAATTTGAACTCTACAGTATGACAGTACTATTTACATATTTGGTATCTCTCTGA 2160  
QY 2161 ATTCTTTTACGATGTTGATGTCAACATTTAGTCTGTCGGATGGCTCTCAAAATGGCAGA 2220  
Db 2161 ATTCTTTTACGATGTTGATGTCAACATTTAGTCTGTCGGATGGCTCTCAAAATGGCAGA 2220  
QY 2221 GGAAAGGATAATGGAAGAGCAGCAGAAAGGCGGTAGTAGTAAAAAACAAGAAAAA 2280  
Db 2221 GGAAAGGATAATGGAAGAGCAGCAGAAAGGCGGTAGTAGTAAAAAACAAGAAAAA 2280  
QY 2281 GAAAGTTCGCCCATTTGAGCCGAGAGATCAATGAGCCAGCATATCAGAACATGTGTGC 2340  
Db 2281 GAAAGTTCGCCCATTTGAGCCGAGAGATCAATGAGCCAGCATATCAGAACATGTGTGC 2340  
QY 2341 TGAATGTTTAAAAACCATGGTAGCATTTGACATGGACGGCAAGTACCTTAAACCGAAGTT 2400

2341 TGGAAATGTTAAACCATGGTAGCATTTGACATGGACGCAAGTACGTAAACCGAAGTT 2400  
2401 TGAGCTTGATAGTGAACAAGTTCCGATGATGAACACAGGTTTGCTCCATTCAACAGTGTGAT 2460  
2401 TGAGCTTGATAGTGAACAAGTTCCGATGATGAACACAGGTTTGCTCCATTCAACAGTGTGAT 2460  
2461 GACCCCGCCCGCCAGTGCATCTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAG 2520  
2461 GACCCCGCCCGCCAGTGCATCTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAG 2520  
2521 CCCTCCTCCTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCAAA 2580  
2521 CCCTCCTCCTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCAAA 2580  
2581 AATGATATTGGAATAATATCTCTAACCCTGACCTAGGTTAATAAGATTTTAAAGTTGC 2640  
2581 AATGATATTGGAATAATATCTCTAACCCTGACCTAGGTTAATAAGATTTTAAAGTTGC 2640  
2641 CAAACCCCACTTTGCTGTTATGAAGTTATGAGGAGGAGGACACAAAGGAATCTAAAGT 2700  
2641 CAAACCCCACTTTGCTGTTATGAAGTTATGAGGAGGAGGACACAAAGGAATCTAAAGT 2700  
2701 TCCTCCTGAATTTGATTTCTCTGCTCATATAATATTTCTGTTGTGAACCTTTGTTGAGA 2760  
2701 TCCTCCTGAATTTGATTTCTCTGCTCATATAATATTTCTGTTGTGAACCTTTGTTGAGA 2760  
2761 GAGACTGGGAGGTGGCCATAAAGGGGAGAGTCTTTCTTCAACCCCACTTTAGAGGG 2820  
2761 GAGACTGGGAGGTGGCCATAAAGGGGAGAGTCTTTCTTCAACCCCACTTTAGAGGG 2820  
2821 CACATCACCAAGGCTCCACATCACCGGGAAGTGAAGTATTTCTTGGGTAAACCTCAFTA 2880  
2821 CACATCACCAAGGCTCCACATCACCGGGAAGTGAAGTATTTCTTGGGTAAACCTCAFTA 2880  
2881 TAAGGAATACCTTTAGTTTACAGCCTTATATGACATGAATGAATAAAGTCTGTTTAAAG 2940  
2881 TAAGGAATACCTTTAGTTTACAGCCTTATATGACATGAATGAATAAAGTCTGTTTAAAG 2940  
2941 TGGTTTATTTATGTTCCATGGAAGAACTGGTCTTATGATGATGATGATGATGATGAT 3000  
2941 TGGTTTATTTATGTTCCATGGAAGAACTGGTCTTATGATGATGATGATGATGATGAT 3000  
3001 GGTTTTATTTATGATTTAATCAACAAATCAATTTTATGATGATGATGATGATGATGAT 3060  
3001 GGTTTTATTTATGATTTAATCAACAAATCAATTTTATGATGATGATGATGATGATGAT 3060  
3061 TTATAAAGGTTTATAAATTTCTTGCACAAAAA 3096  
3061 TTATAAAGGTTTATAAATTTCTTGCACAAAAA 3096

RESULT 2

US-09-014-353-20725  
Sequence 20725, Application US/09814353  
Publication No. US20030165831A1  
GENERAL INFORMATION:  
APPLICANT: Lee, John  
APPLICANT: Thompson, Pamela  
APPLICANT: Lillie, James  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF OVARIAN CANCER  
FILE REFERENCE: MRI-006B  
CURRENT APPLICATION NUMBER: US/09/814,353  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: US 60/191,031  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: US 60/207,124  
PRIOR FILING DATE: 2000-05-25  
PRIOR APPLICATION NUMBER: US 60/211,940  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: US 60/216,820

Not Prior Art  
Provisional  
App do not  
disclose sequence  
"20725"

PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: US 60/220,661  
PRIOR FILING DATE: 2000-07-25  
PRIOR APPLICATION NUMBER: US 60/257,672  
PRIOR FILING DATE: 2000-12-21  
NUMBER OF SEQ ID NOS: 22037  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 20725  
LENGTH: 2610  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1, 2, 3, 4  
OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-20725

Query Match 72.0%; Score 2228.4; DB 12; Length 2610;  
Best Local Similarity 97.9%; Pred. No. 0;  
Matches 2299; Conservative 0; Mismatches 21; Indels 28; Gaps 3;  
QY 426 CGCATGCGTGCACGCTGCCGCTGCGGCTGGGCTGAGAGGGAGGGGGCGGCGGCGGCGG 485  
Db 2610 CGCATGCGTGCACGCTGCCGCTGCGGCTGGGCTGAGAGGGAGGGGGCGGCGGCGGCGG 2551  
QY 486 GCGGCGCTGCTTATTTCCGTGTCGCGACAGTCCGTGCGGCGGCGGCGGCGGCGGCGG 545  
Db 2550 GCGGCGCTGCTTATTTCCGTGTCGCGACAGTCCGTGCGGCGGCGGCGGCGGCGGCGG 2491  
QY 546 AGTAGGCATAAATGGTTATGAAAGTCTCTGTAGATGATGACGATTCAGGATGGAGCTCAG 605  
Db 2490 AGTAGGCATAAATGGTTATGAAAGTCTCTGTAGATGATGACGATTCAGGATGGAGCTCAG 2431  
QY 606 TATGCCAGAAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTAACCAAGATTTTGA 665  
Db 2430 TATGCCAGAAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTAACCAAGATTTTGA 2371  
QY 666 AGAAGCTTGTGCGAGAAATTAAGTTGGGAGAACTACTTCAATGATAAGCTATTTGGTCTTT 725  
Db 2370 AGAAGCTTGTGCGAGAAATTAAGTTGGGAGAACTACTTCAATGATAAGCTATTTGGTCTTT 2311  
QY 726 TGAAGCCATGCTGCTATTTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGA 785  
Db 2310 TGAAGCCATGCTGCTATTTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGA 2251  
QY 786 CCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGATGGCACTATTAAT 845  
Db 2250 CCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGATGGCACTATTAAT 2191  
QY 846 TAAAGATCTCACCTTGCTGAACTGATAGGATTAAGATGATGATGATGATGATGATGAT 905  
Db 2190 TAAAGATCTCACCTTGCTGAACTGATAGGATTAAGATGATGATGATGATGATGATGAT 2131  
QY 906 AACGTGTTAGAGGCGCATTCATCGGCACAGACAGTATTTACGTGCTTACATTCATAA 965  
Db 2130 AACGTGTTAGAGGCGCATTCATCGGCACAGACAGTATTTACGTGCTTACATTCATAA 2071  
QY 966 TCCAGACTTTATAGAAAGTCTGCTATGAAGGCTTTTCTCTGGAATCTTGAAATCTG 1025  
Db 2070 TCCAGACTTTATAGAAAGTCTGCTATGAAGGCTTTTCTCTGGAATCTTGAAATCTG 2011  
QY 1026 TGACATTGCAAGGAAAAAGTAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTC 1085  
Db 2010 TGACATTGCAAGGAAAAAGTAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTC 1951  
QY 1086 AATGACTTATGATTTAAATGGCTTAACAGTGTGACAGATCTTCAGTTACAGGCTGCT 1145  
Db 1950 AATGACTTATGATTTAAATGGCTTAACAGTGTGACAGATCTTCAGTTACAGGCTGCT 1891  
QY 1146 AAAAGATGTGGAGGATGACATGCAACAGAGTAAAGATCTGCAAGTCCGACAGGAGA 1205  
Db 1890 AAAAGATGTGGAGGATGACATGCAACAGAGTAAAGATCTGCAAGTCCGACAGGAGA 1831

QY 1206 AGAAAGAGATCCAGAGTTGAAGTAAACACCAACCAATGTTTAGCAGTATTCAGCAGAGT 1265  
Db 1830 AGAAAGAGATCCAGAGTTGAAGTAAACACCAACCAATGTTTAGCAGTATTCAGCAGAGT 1771  
QY 1266 GAAATTTACTCGTGTGTACTGACAGTGCTTTATAGCCTTTTACTAAGAAAGAGACCAAGTGC 1325  
Db 1770 GAAATTTACTCGTGTGTACTGACAGTGCTTTATAGCCTTTTACTAAGAAAGAGACCAAGTGC 1711  
QY 1326 TGTTGCAAGACTCAAAATTTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCTATAA 1385  
Db 1710 TGTTGCAAGACTCAAAATTTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCTATAA 1651  
QY 1386 TTCAATTCATCATGGCATCCAGGCCCCAGAAATGATACATAAAGGAGATCATCCAATTAT 1445  
Db 1650 TTCAATTCATCATGGCATCCAGGCCCCAGAAATGATACATAAAGGAGATCATCCAATTAT 1591  
QY 1446 GATGGTTTTGAACCCCTTGTGAACAGGAGGCTACTTCCACCTACCTTCCCTCGATATGC 1505  
Db 1590 GATGGTTTTGAACCCCTTGTGAACAGGAGGCTACTTCCACCTACCTTCCCTCGATATGC 1531  
QY 1506 AAAAATAATTAAGGGAAGAAATGGTGAACCTATTTTGAAGATTAATAGATAGATAA 1565  
Db 1530 AAAAATAATTAAGGGAAGAAATGGTGAACCTATTTTGAAGATTAATAGATAGATAA 1471  
QY 1566 AACTGCTGTGAGTTGTGAATTTAAACAAATTTACATTTGTATCTCTGATTTTCTGTGA 1625  
Db 1470 AACTGCTGTGAGTTGTGAATTTAAACAAATTTACATTTGTATCTCTGATTTTCTGTGA 1411  
QY 1626 ATTTAGTGAACAGTCAACATGTGTCTTTTCAAGATCTCTGTACAAACCACTTCTCTGGT 1685  
Db 1410 ATTTAGTGAACAGTCAACATGTGTCTTTTCAAGATCTCTGTACAAACCACTTCTCTGGT 1351  
QY 1686 GGATAACAAAAAGGTCTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACTTCG 1745  
Db 1350 GGATAACAAAAAGGTCTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACTTCG 1291  
QY 1746 GTCTTTTGTGATCCTCCGAGTGTCTTCCCAAGTGTCTACCTATATAATATCACCAG 1805  
Db 1290 GTCTTTTGTGATCCTCCGAGTGTCTTCCCAAGTGTCTACCTATATAATATCACCAG 1233  
QY 1806 GCTAAGGACTGTATCGACTCCTTTGTACTCACTGTGTCTGGCCATTTCTGTAGTCTTATT 1865  
Db 1232 GCTAAGGACTGTATCGACTCCTTTGTACTCACTGTGTCTGGCCATTTCTGTAGTCTTATT 1173  
QY 1866 CAGATCCATGACATAAACAGGGCTCGACAGAGAGATAAGCTTGTCTATATTTCTTGAGGAA 1925  
Db 1172 CAGATCCATGACATAAACAGGGCTCGACAGAGAGATAAGCTTGTCTATATTTCTTGAGGAA 1113  
QY 1926 TTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTTAAATAGGCAGAGAAGTTGATG 1985  
Db 1112 TTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTTAAATAGGCAGAGAAGTTGATG 1076  
QY 1986 CAGCGCTTCACACCATGCTGTGTGAACACAGGAACCCCAAGGCAACATTTGGCCTGTGTAG 2045  
Db 1075 CAGCGCTTCACACCATGCTGTGTGAACACAGGAACCCCAAGGCAACATTTGGCCTGTGTAG 1016  
QY 2046 GTACCTGGGTCTTTACCATAAACCTTCGCATTATGATACAGTACCTTCTAAGTGGCTTTG 2105  
Db 1015 GTACCTGGGTCTTTACCATAAACCTTCGCATTATGATACAGTACCTTCTAAGTGGCTTTG 956  
QY 2106 AATTGGAACTCTACAGTATGACAGAGTACTATACATATATTTGGTATCTCTCTGAATTC 2165  
Db 955 AATTGGAACTCTACAGTATGACAGAGTACTATACATATATTTGGTATCTCTCTGAATTC 896  
QY 2166 TTTACGCATGGTTGATGTCAACATTTGAGTCTGTCGCGATGGCTCTCAATGGCAGAGGAA 2225  
Db 895 TTTACGCATGGTTGATGTCAACATTTGAGTCTGTCGCGATGGCTCTCAATGGCAGAGGAA 836  
QY 2226 GGATAATGGAAGAGCAGAGAAAGCCCGTAGTAGTAAAAAACAAGAAAAAGAAAG 2285  
Db 835 GGATAATGGAAGAGCAGAGAAAGCCCGTAGTAGTAAAAAACAAGAAAAAGAAAG 776  
QY 2286 TTCGCCCATTTGAGCCGAGAGATCAAAATGAGCCAGCATATCAGAAATGTGTGCTGGAA 2345

Db 775 TTCGCCCATTTGAGCCGAGAGATCACAATGAGCCCAAGCATATCAGAAATGTGTGCTGGA 716  
QY 2346 TGTTTAAACCAATGGTAGCATTTTACATGACGCGCAAGTAGTAAACCGAAGTTTGAGC 2405  
Db 715 TGTTTAAACCAATGGTAGCATTTTACATGACGCGCAAGTAGTAAACCGAAGTTTGAGC 656  
QY 2406 TTGATAGTGAACAAAGTTCCGGTATGAACACAGAGTTTGTCTCCATTCAACAGTGTGATGACCC 2465  
Db 655 TTGATAGTGAACAAAGTTCCGGTATGAACACAGAGTTTGTCTCCATTCAACAGTGTGATGACCC 596  
QY 2466 CGCCGCCAGTGCATCTACTTACAGTTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTC 2525  
Db 595 CGCCGCCAGTGCATCTACTTACAGTTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTC 536  
QY 2526 CTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACCTTTCAACAGGCAAAATGA 2585  
Db 535 CTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACCTTTCAACAGGCAAAATGA 476  
QY 2586 TATTGGAAATATCTCTAACCCTGACCTAGTAAGCACCTTTCAACAGGCAAAATGA 2645  
Db 475 TATTGGAAATATCTCTAACCCTGACCTAGTAAGCACCTTTCAACAGGCAAAATGA 416  
QY 2646 CCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAGGATCTAAAGTTCTC 2705  
Db 415 CCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAGGATCTAAAGTTCTC 356  
QY 2706 CTGAATTTGATTTCTCTGCTCATATAATTTTCTCTGTTGTGAAACTTGTGAGAGAGAC 2765  
Db 355 CTGAATTTGATTTCTCTGCTCATATAATTTTCTCTGTTGTGAAACTTGTGAGAGAGACTG 299  
QY 2766 TGGGAGG 2773  
Db 298 GGAGTGG 291

RESULT 3  
US-09-814-353-15927/c  
; Sequence 15927, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15927  
; LENGTH: 745  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 571  
; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-15927

Query Match 18.1%; Score 560.8; DB 12; Length 745;  
Best Local Similarity 94.8%; Pred. No. 2.2e-141;  
Matches 621; Conservative 0; Mismatches 8; Indels 26; Gaps 3;  
QY 1399 GGCATCCAGGCCAGGATGATACATAAAGAGATCATCCAAATATGATGGTTTGA 1458  
Db |||||||  
QY 1459 CCCCTGTGAACAGGCTACTTCCACCTACCTCCCTCGATATGCAAAATAATATA 1518  
Db |||||||  
QY 1519 AGGGAAGAAATGGTGAATATTTTGAAGATTAATAGATAGAAATAAATAATA 1578  
Db |||||||  
QY 1579 GTTGTGAATTAACAAATTAATGATCTCTGA -TTTTTCTGTGAATTAATGA 1637  
Db |||||||  
QY 1638 GTACCATGTGTTCTTCAAGATCTCTGTTACAAACCACTTCTGTTGATGA 1697  
Db |||||||  
QY 1698 GTTGTGAATTAACAAATTAATGATCTCTGTTACAAACCACTTCTGTTGATGA 1757  
Db |||||||  
QY 1758 ATCTCCGAGTGTCTTCCCAAGTGTCTACCTATATAATCAACAGGCTAAGG 1817  
Db |||||||  
QY 1818 ATCGACTCTTGTGTTACTACTGTGTTCCGCCATCTGTAGTCTTATCAGATCC 1877  
Db |||||||  
QY 1878 CATAACAGGCTCGACAGAGATAAGCTTGGTCAATCTTGTAGGAAATTTGCC 1937  
Db |||||||  
QY 1938 CAGGATGAGTTTATGACATTTTATTTAATAGGAGAGAGTTGATGAGCGCTT 1997  
Db |||||||  
QY 1998 CCATGCTGTGAAACAGGACCCCAAGGCAACATTTGGCCTGTTTAGTACCTG 2052  
Db |||||||  
QY 2052 CCATGCTGTGAAACAGGACCCCAAGGCAACATTTGGCCTGTTTAGTACCTG 2110  
Db |||||||

## RESULT 4

US-09-814-353-16155  
; Sequence 16155, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661

; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16155  
; LENGTH: 610  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-814-353-16155

Query Match 16.5%; Score 509.4; DB 12; Length 610;  
Best Local Similarity 99.8%; Pred. No. 1.8e-127;  
Matches 510; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1184 GTACTCGAAGTCGACAGGAGAGAGATCCAGAAAGTTGAACACTAGAACCAACAAT 1243  
Db |||||||  
QY 1244 GTTTCAGAGTATTCAGCAGAGTGAATTTACTCGTGTGTTACTGACAGTGTATAGCCT 1303  
Db |||||||  
QY 1304 TTACTAAGAAAGAGACCAAGTGTGTCAGAAAGCTCAAAAATTTGATGGTTCAAGCAGCAG 1363  
Db |||||||  
QY 1364 ATCTTCTTCTGCCATTCATAATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1423  
Db |||||||  
QY 1424 CAAAGGAGATCATCCAATTAATGATGGGTTTGTGAACCCCTTGTGAACCCAGAGGCTACTTC 1483  
Db |||||||  
QY 1484 CACCTACCTTCCCTCGATATGCAAAAATAATTAAGGGAAGAAATGGTGAACCTATTTTG 1543  
Db |||||||  
QY 1544 CAAGATTAATAGATAGAAATAAAACTGTCTGTGAGGTTGTGAATTTAACAATTTACATT 1603  
Db |||||||  
QY 1604 GTATCCTGGATTTTCTGTGAATTTAGTGAACAGTCACTGTGTCTTTCAAGATCTC 1663  
Db |||||||  
QY 1664 TGTACAAACCACTTCTCGTGGTGAATAACAA 1694  
Db |||||||  
QY 1694 TGTACAAACCACTTCTCGTGGTGAATAACAA 610

## RESULT 5

US-09-814-353-3218/c  
; Sequence 3218, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820



Db 87 GCGCTTCACACCATGCTGTTGAACACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGGT 28  
QY 2048 ACCTG 2052  
| | | | |  
Db 27 ACCTG 23

## RESULT 7

US-09-918-995-20914  
; Sequence 20914, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 20914  
; LENGTH: 482  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(482)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-20914

Query Match 14.6%; Score 451.2; DB 11; Length 482;  
Best Local Similarity 99.3%; Pred. No. 9.9e-112;  
Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 491 CGTCGTTATTTCCGTCGTCGACAGTGGTGGCGGCGGGTGACCAACGGGAGAGTAG 550  
| | | | |  
Db 26 CGAGCTTATTTCCGTCGTCGACAGTGGTGGCGGCGGGTGACCAACGGGAGAGTAG 85  
| | | | |  
QY 551 GCATAATGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGGAGCTCAGTATGC 610  
| | | | |  
Db 86 GCATAATGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGGAGCTCAGTATGC 145  
| | | | |  
QY 611 CAGAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTAACCAAGATTTTGAAGAAG 670  
| | | | |  
Db 146 CAGAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTAACCAAGATTTTGAAGAAG 205  
| | | | |  
QY 671 CTTGTCGAGAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTGTCCTTTTGAAG 730  
| | | | |  
Db 206 CTTGTCGAGAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTGTCCTTTTGAAG 265  
| | | | |  
QY 731 CCATGTCGTCTATTGAAATGATGGATCCCAAGATGGATGCTGSCATGATTGGAAACCAAG 790  
| | | | |  
Db 266 CCATGTCGTCTATTGAAATGATGGATCCCAAGATGGATGCTGSCATGATTGGAAACCAAG 325  
| | | | |  
QY 791 TTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGSCACTATTAAATTAAG 850  
| | | | |  
Db 326 TTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGSCACTATTAAATTAAG 385  
| | | | |  
QY 851 ATCTCACCTTGCCTGAACTGATAGGATTAATGATACATGTTTGTCTGTTGATAACGT 910  
| | | | |  
Db 386 ATCTCACCTTGCCTGAACTGATAGGATTAATGATACATGTTTGTCTGTTGATAACGT 445  
| | | | |  
QY 911 GGTAGAAAGGCCATTCACTGGCACAGACAGTATTTA 946  
| | | | |  
Db 446 GGTAGAAAGGCCATTCACTGGCACAGACAGTATTTA 481  
| | | | |

## RESULT 8

US-09-814-353-3455  
; Sequence 3455, Application US/09814353  
; Publication No. US20030165831A1

; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3455  
; LENGTH: 434  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-814-353-3455

Query Match 14.0%; Score 434; DB 12; Length 434;  
Best Local Similarity 100.0%; Pred. No. 4.3e-107;  
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1186 ACTCGAAGTCGACAAAGGAGAGAGAGATCCAGAGATTGAACCTAGAACACCAACAATGT 1245  
| | | | |  
Db 1 ACTCGAAGTCGACAAAGGAGAGAGAGATCCAGAGATTGAACCTAGAACACCAACAATGT 60  
| | | | |  
QY 1246 TTAGCAGTATTCAGCAGAGTGAAATTTACTCGTGTGTACTGACAGTGCTTATAGCCTTT 1305  
| | | | |  
Db 61 TTAGCAGTATTCAGCAGAGTGAAATTTACTCGTGTGTACTGACAGTGCTTATAGCCTTT 120  
| | | | |  
QY 1306 ACTAAGAAAGAGACCAGTGCTGTTCAGAAAGCTCAAAAATTTGATGTTCAAGCAGCAGAT 1365  
| | | | |  
Db 121 ACTAAGAAAGAGACCAGTGCTGTTCAGAAAGCTCAAAAATTTGATGTTCAAGCAGCAGAT 180  
| | | | |  
QY 1366 CTTCTTTCTGCCATTTCATAATTTCATTCATGTCATCATGGCCATCCAGGCCAGAAATGATACTACA 1425  
| | | | |  
Db 181 CTTCTTTCTGCCATTTCATAATTTCATTCATGTCATCATGGCCATCCAGGCCAGAAATGATACTACA 240  
| | | | |  
QY 1426 AAAGGAGATCATCCAATTATGATGGGTTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCA 1485  
| | | | |  
Db 241 AAAGGAGATCATCCAATTATGATGGGTTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCA 300  
| | | | |  
QY 1486 CCTACCTTCCCTCGATATGCAAAAATAATTAAAGGGAAGAAATGGTGAACCTATTTTGA 1545  
| | | | |  
Db 301 CCTACCTTCCCTCGATATGCAAAAATAATTAAAGGGAAGAAATGGTGAACCTATTTTGA 360  
| | | | |  
QY 1546 AGATTAAATAGATAGAAATAAAACTGTCTGTGAGGTTGTGAATTTAACAAATTTACATTGT 1605  
| | | | |  
Db 361 AGATTAAATAGATAGAAATAAAACTGTCTGTGAGGTTGTGAATTTAACAAATTTACATTGT 420  
| | | | |  
QY 1606 ATCCTGGATTTTTT 1619  
| | | | |  
Db 421 ATCCTGGATTTTTT 434  
| | | | |

## RESULT 9

US-09-814-353-9771  
; Sequence 9771, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John

; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9771  
; LENGTH: 434  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-814-353-9771

Query Match 14.0%; Score 434; DB 12; Length 434;  
Best Local Similarity 100.0%; Pred. No. 4.3e-107;  
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1186 ACTCGAAGTCGACAAAGGAGAGAAAGAGATCCAGAAAGTTGAACTAGAACACCAACAATGT 1245  
Db 1 ACTCGAAGTCGACAAAGGAGAGAAAGAGATCCAGAAAGTTGAACTAGAACACCAACAATGT 60  
QY 1246 TTAGCAGTATTCAGCAGAGTGAAATTTACTCGTGTGTACTGACAGTGCTTATAGCCTTT 1305  
Db 61 TTAGCAGTATTCAGCAGAGTGAAATTTACTCGTGTGTACTGACAGTGCTTATAGCCTTT 120  
QY 1306 ACTAAGAAAGAGACACAGTGCTGTTGCAGAAAGCTCAAAAATTGATGTTCAAGCAGCAGAT 1365  
Db 121 ACTAAGAAAGAGACACAGTGCTGTTGCAGAAAGCTCAAAAATTGATGTTCAAGCAGCAGAT 180  
QY 1366 CTTCTTTCTGCCATTTCATAATTCATTGTCATGCGATCCAGGCCCGAGAAATGATACATA 1425  
Db 181 CTTCTTTCTGCCATTTCATAATTCATTGTCATGCGATCCAGGCCCGAGAAATGATACATA 240  
QY 1426 AAAGGAGATCATCCAAATATGATGSGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCA 1485  
Db 241 AAAGGAGATCATCCAAATATGATGSGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCA 300  
QY 1486 CCTACCTTCCCTCGATATGCAAAAATAATTAAGGAGAAATGGTGAACCTATTTTGCA 1545  
Db 301 CCTACCTTCCCTCGATATGCAAAAATAATTAAGGAGAAATGGTGAACCTATTTTGCA 360  
QY 1546 AGATTAATAGATAGATAAATACTGCTGAGGTTGTGAATTTAACAATAATTACATTGT 1605  
Db 361 AGATTAATAGATAGATAAATACTGCTGAGGTTGTGAATTTAACAATAATTACATTGT 420  
QY 1606 ATCCTGGATTTTTT 1619  
Db 421 ATCCTGGATTTTTT 434

RESULT 10  
US-09-983-965-210  
; Sequence 210, Application US/09983965  
; Patent No. US20020137160A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 09/465,231  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: US 60/113,678  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 210  
; LENGTH: 398  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 25-BOVMS1-018-Q1-E1-G1  
; US-09-983-965-210

Query Match 11.3%; Score 350.8; DB 10; Length 398;  
Best Local Similarity 94.3%; Pred. No. 1.7e-84;  
Matches 364; Conservative 0; Mismatches 22; Indels 0; Gaps 0;  
QY 2298 GCCGAGAGATCACAATGAGCCCAAGCATATCAGAACATGTGCTGGAATGTTTAAACCA 2357  
Db 13 GCCGAGAGATCACAATGAGCCCAAGCATATCAGAACATGTGCTGGAATGTTTAAACCA 72  
QY 2358 TGGTAGCATTTGACATGAGCCGCAAGTACGTAAACCGAAGTTTGAGCTTGATAGTGAAC 2417  
Db 73 TGGGAGCTTTGACATGAGTGGCAAGTACGAAACCCCAAGTTTGAGCTTGATAGTGAAC 132  
QY 2418 AAGTTCCGATGAACACAGAGTTTGTCTCCATTCAACAGTGTGATGACCCCGCCAGTGC 2477  
Db 133 AAGTTCCGATGAACACAGAGTTTGTCTCCATTCAACAGTGTGATGACACACCACAGTGC 192  
QY 2478 ACTACTTACAGTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTCCTCAGTCTC 2537  
Db 193 ACTATCTGAGTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTCCTCAGTCTC 252  
QY 2538 CTGAACCTGTATGGCAGCTAGTAAGCACTTTCAACAGGCAAAATGATATTGGAATA 2597  
Db 253 CAGAACTGTATGGCAGCTAGTAAGCACTTTCAACAGGCAAAATGATATTGGAATA 312  
QY 2598 TTCTTAACCCGACCATGAGTTTAATAGAAATTTAAAGGTTGCCAAACCCCACTTTGTGG 2657  
Db 313 TCCCAAAACCCAGACCATGAGTCAATAGAAATTTAAAGGTTGCCAAACCCCACTTTGTGG 372  
QY 2658 TTATGAAGTTTATGGCAGGAGACAC 2683  
Db 373 TTATGAAGTTTATGGCAGGAGACAC 398

RESULT 11  
US-09-783-590-5693  
; Sequence 5693, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16.2C1  
; CURRENT APPLICATION NUMBER: US/09/783,590  
; CURRENT FILING DATE: 2000-02-15  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1995-04-12  
; PRIOR APPLICATION NUMBER: 08/346,731  
; PRIOR FILING DATE: 1994-11-21  
; NUMBER OF SEQ ID NOS: 12485  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5693



```

; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 323193
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-323193

Query Match      5.9%; Score 182; DB 13; Length 489;
Best Local Similarity 100.0%; Pred. No. 1.6e-38;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1431 AGATCATCCAATTATGATGGGTTTGAACCCCTTGTGAACGAGGCTACTTCCACCTAC 1490
Db      401 AGATCATCCAATTATGATGGGTTTGAACCCCTTGTGAACGAGGCTACTTCCACCTAC 342

QY      1491 CTTCCCTCGATATGCAAAAATAAATAAAGGGAAGAAATGTTGAACTATTTTGCAAGATT 1550
Db      341 CTTCCCTCGATATGCAAAAATAAATAAAGGGAAGAAATGTTGAACTATTTTGCAAGATT 282

QY      1551 AATAGATAGAATAAAAACCTGCTGTGAGGTTGTGAATTTAACAAATTACATTGTATCCT 1610
Db      281 AATAGATAGAATAAAAACCTGCTGTGAGGTTGTGAATTTAACAAATTACATTGTATCCT 222

QY      1611 GG 1612
Db      221 GG 220

RESULT 14
US-10-001-857-41/c
; Sequence 41, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prob
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 151
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-857-41

Query Match      4.5%; Score 140.4; DB 13; Length 151;
Best Local Similarity 99.3%; Pred. No. 1.3e-27;
Matches 142; Conservative 0; Mismatches 2; Indels 23; Gaps 1;

QY      1910 TCATATTCTTGAGGAATTGCCACCTTGCCAGGATGAGTTTATGACATTTTATTTTATAG 1969
Db      205 TCATATTCTTGAGGAATTGCCACCTTGCCAGGATG-----AG 169

QY      1970 GCAGAGAAGGTTTGATGCAGCGCTTCACACCATGCTGTTGAAAACAGGAACCCCAAGGCAA 2029
Db      168 GCAGAGAAGGTTTGATGCAGCGCTTCACACCATGCTGTTGAAAACAGGAACCCCAAGGCAA 109

QY      2030 CATTGGCCTGTTTAGGTACCTGGGTC 2056
Db      108 CATTGGCCTGTTTAGGTACCTCGGCC 82

Search completed: November 25, 2003, 03:08:51
Job time : 938 secs
```

```

Matches 141; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1911 CATATTCTTGAGGAATTGCCACCTTGCCAGGATGAGTTTATGACATTTTATTTAATAGG 1970
Db      151 CATATTCTTGAGGAATTGCCACCTTGCCAGGATGAGTTTATGACATTTTATTTAATAGG 92

QY      1971 CAGAGAAGGTTGATGCAGCGCTTCACACCATGCTGTTGAAAACAGGAACCCCAAGGCAAC 2030
Db      91 CAGAGAAGGTTGATGCAGCGCTTCACACCATGCTGTTGAAAACAGGAACCCCAAGGCAAC 32

QY      2031 ATTTGGCCTGTTTAGGTACCTG 2052
Db      31 ATTTGGCCTGTTTAGGTACCTG 10

RESULT 15
US-10-001-857-111/c
; Sequence 111, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 111
; LENGTH: 815
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (206)..(305)
; OTHER INFORMATION: a, c, g or t
US-10-001-857-111

Query Match      2.8%; Score 87.8; DB 13; Length 815;
Best Local Similarity 83.0%; Pred. No. 9.4e-13;
Matches 122; Conservative 0; Mismatches 2; Indels 23; Gaps 1;

QY      1910 TCATATTCTTGAGGAATTGCCACCTTGCCAGGATGAGTTTATGACATTTTATTTTATAG 1969
Db      205 TCATATTCTTGAGGAATTGCCACCTTGCCAGGATG-----AG 169

QY      1970 GCAGAGAAGGTTTGATGCAGCGCTTCACACCATGCTGTTGAAAACAGGAACCCCAAGGCAA 2029
Db      168 GCAGAGAAGGTTTGATGCAGCGCTTCACACCATGCTGTTGAAAACAGGAACCCCAAGGCAA 109

QY      2030 CATTGGCCTGTTTAGGTACCTGGGTC 2056
Db      108 CATTGGCCTGTTTAGGTACCTCGGCC 82

Search completed: November 25, 2003, 03:08:51
Job time : 938 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2003, 21:45:44 ; Search time 779 Seconds  
(without alignments)  
10728.447 Million cell updates/sec

Title: US-10-001-857-42  
Perfect score: 3096  
Sequence: 1 ttctcagaaactccagg.....atttcttgacaaaaaaaaa 3096

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 19Jun03:\*  
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3096	100.0	3096	24	ABQ75303 Human lung specifi
C	2577.8	83.3	2668	23	ABV23940 Human prostate exp
C	2577.8	83.3	2668	23	ABV29823 Human prostate exp
4	2417.2	78.1	2488	24	AAL49929 Human molecule for
5	717.4	23.2	719	24	ABS51309 cDNA encoding huma
6	350.8	11.3	398	25	ABX50281 Bovine EST associa
7	343.8	11.1	799	24	ABQ51658 Oligonucleotide fo
C	343.8	11.1	799	24	ABQ51659 Oligonucleotide fo

C	9	292.4	9.4	356	23	ABV15331	Human prostate exp
C	10	288.8	9.3	439	23	ABV36125	Human prostate exp
C	11	288.8	9.3	439	23	ABV45182	Human prostate exp
C	12	274.2	8.9	799	24	ABQ51660	Oligonucleotide fo
C	13	274.2	8.9	799	24	ABQ51661	Oligonucleotide fo
C	14	230.2	7.4	303	23	ABV06162	Human prostate exp
C	15	217	7.0	2515	23	ABL07143	Drosophila melanog
C	16	140.4	4.5	151	24	ABQ75302	Human lung specifi
C	17	90.6	2.9	3401	23	ABL05906	Drosophila melanog
C	18	90.6	2.9	4847	23	ABL07142	Drosophila melanog
C	19	87.8	2.8	815	24	ABQ75372	Human lung specifi
C	20	60	1.9	60	24	ABN43747	Human spliced tran
C	21	58.6	1.9	600	24	ABQ52496	Oligonucleotide fo
C	22	58.6	1.9	600	24	ABQ52497	Oligonucleotide fo
C	23	56.4	1.8	320	21	AAA38183	Primer used in the
C	24	56	1.8	451	24	ABT10409	Human breast cance
C	25	56	1.8	1286	24	ABI99656	Mouse ischaemic co
C	26	55.4	1.8	365	23	ABV54856	Human prostate exp
C	27	55.2	1.8	320	21	AAA38186	Primer used in the
C	28	54.8	1.8	320	21	AAA38185	Primer used in the
C	29	54.6	1.8	840	24	ABQ35494	Oligonucleotide fo
C	30	54.6	1.8	840	24	ABQ35495	Oligonucleotide fo
C	31	54.6	1.8	1416	25	ABZ20967	Animal test kit ma
C	32	54.6	1.8	12733	24	ABK98631	Vector pPEPF14 con
C	33	54.6	1.8	12739	24	ABK98592	Vector pPEPF1 cont
C	34	54.2	1.8	318	21	AAA38184	Primer used in the
C	35	54.2	1.8	4403765	22	AAI99683	Mycobacterium tube
C	36	54.2	1.8	4411529	22	AAI99682	Mycobacterium tube
C	37	53.8	1.7	2188	20	AAZ77506	Human ovarian tumo
C	38	53.4	1.7	1337	20	AAZ17263	Human gene express
C	39	53.2	1.7	2561	22	AAH26500	Rabbit low density
C	40	53	1.7	434	22	AAI84343	Human polynucleoti
C	41	52.8	1.7	344	22	AAI82007	Human polynucleoti
C	42	52.2	1.7	1327	24	ABQ68452	Listeria monocytog
C	43	51.6	1.7	1000	21	AAA02484	Human colon cancer
C	44	51.4	1.7	237	21	AAA81807	N. meningitidis pa
C	45	51	1.6	712	24	ABQ40858	Oligonucleotide fo

ALIGNMENTS

RESULT 1

ABQ75303  
ID ABQ75303 standard; cDNA; 3096 BP.

XX ABQ75303;

AC ABQ75303;

XX 05-NOV-2002 (first entry)

DE Human lung specific nucleic acid sequence SEQ ID NO:42.  
XX Human; lung; lung specific nucleic acid; LSNA; lung specific protein;  
KW LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;  
KW squamous cell carcinoma; gene; chromosome 9; ss.

XX Homo sapiens.

OS WO200264788-A2.

PN 22-AUG-2002.

XX 20-NOV-2001; 2001WO-US45080.

XX 20-NOV-2000; 2000US-252054P.

XX (DIAD-) DIADEXUS INC.

PI Macina RA, Recipon H, Chen S, Sun Y, Liu C;

XX WPI; 2002-657601/70.

XX New lung specific nucleic acid useful in gene therapy or as vaccines

PT for treating lung cancer (e.g. squamous cell carcinoma) or  
PT non-cancerous lung diseases, as well as for diagnosing, monitoring or  
PT staging these diseases -  
XX  
  
PS Claim 1; Page 168-169; 282pp; English.  
XX  
  
CC The present invention describes an isolated lung specific nucleic acid  
CC (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid  
CC sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;  
CC (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp),  
CC given in ABQ75262 to ABQ75376; (c) selectively hybridises to (a) or (b);  
CC or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific  
CC protein (LSP) sequences have cytostatic activity and can be used in gene  
CC therapy and vaccines. LSNAs and LSPs are useful for diagnosing and  
CC monitoring the presence and metastases of lung cancer in a patient. An  
CC antibody that specifically binds to an LSP can be used for determining  
CC the presence of an LSP in a sample, as well as for treating a patient  
CC with lung cancer, particularly by inducing an immune response against  
CC the lung cancer cell expressing the LSNAs or LSPs. In particular, these  
CC LSNAs and LSPs are useful for identifying, diagnosing, monitoring,  
CC staging, imaging and treating lung cancer (e.g. squamous cell carcinoma)  
CC and non-cancerous disease states in lung.  
XX  
SQ Sequence 3096 BP; 914 A; 606 C; 749 G; 827 T; 0 other;

Query Match 100.0%; Score 3096; DB 24; Length 3096;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3096; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTCTACGAAACTCCAGGCGTGTATAGGAAACATAAATCCGTTGTTCAGGCAGCAGTA 60  
DB 1 TTCTCTACGAAACTCCAGGCGTGTATAGGAAACATAAATCCGTTGTTCAGGCAGCAGTA 60  
  
QY 61 GCACGCTGTTGCTCTCGGAGCTTGCGTCTCGTCTCGCAACCACCTAAGGTCTACG 120  
DB 61 GCACGCTGTTGCTCTCGGAGCTTGCGTCTCGTCTCGCAACCACCTAAGGTCTACG 120  
  
QY 121 CAAACCTCCACGGTTTCCTTCGCGCTTCGCGTCACTTCTAAGAAATCCACAGAGGCA 180  
DB 121 CAAACCTCCACGGTTTCCTTCGCGCTTCGCGTCACTTCTAAGAAATCCACAGAGGCA 180  
  
QY 181 GCGCAGACGGGGCGGCTCTGACACTCCGGGCTCCGCCCTCTTCCGGGAACCGCCCACTA 240  
DB 181 GCGCAGACGGGGCGGCTCTGAGACTCCGGGCTCCGCCCTCTTCCGGGAACCGCCCACTA 240  
  
QY 241 CCCAGGACTCCGACAGAGGTTGAAAGATAAATCTCCGGTCTCGCGATCGTCTCTAATC 300  
DB 241 CCCAGGACTCCGACAGAGGTTGAAAGATAAATCTCCGGTCTCGCGATCGTCTCTAATC 300  
  
QY 301 TCGCGAGAAAGAGAGCGGCCCATCGGCCGAACGGAGGCGGTGGCGAGGGGGGTG 360  
DB 301 TCGCGAGAAAGAGAGCGGCCCATCGGCCGAACGGAGGCGGTGGCGAGGGGGGTG 360  
  
QY 361 TGGCCGGGGAGCGGAAGTCCCGGGAGTAAGGGAGAGGGCGGGTTCGCGCTCCCGG 420  
DB 361 TGGCCGGGGAGCGGAAGTCCCGGGAGTAAGGGAGAGGGCGGGTTCGCGCTCCCGG 420  
  
QY 421 GCATACGCATGCGTGCACGCTGCCGTCGGGCTGGGCTGAGAGGGGAGGGCGCGCGG 480  
DB 421 GCATACGCATGCGTGCACGCTGCCGTCGGGCTGGGCTGAGAGGGGAGGGCGCGCGG 480  
  
QY 481 GCCGAGCGCGCTCGTTATTTCCGTGGTCCGGACAGTCCGTGGCGCGCGGTGACCCAG 540  
DB 481 GCCGAGCGCGCTCGTTATTTCCGTGGTCCGGACAGTCCGTGGCGCGCGGTGACCCAG 540  
  
QY 541 GGAGAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATCATGACGATTCAGGATGGGAG 600  
DB 541 GGAGAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATCATGACGATTCAGGATGGGAG 600  
  
QY 601 CTCAGTATGCCAGAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTACCCAAAGAT 660  
DB 601 CTCAGTATGCCAGAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTACCCAAAGAT 660

QY 661 TTTGAAGAAGCTTGTGAGAAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTGGT 720  
DB 661 TTTGAAGAAGCTTGTGAGAAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTGGT 720  
  
QY 721 CTTTGTGAAGCCATGCTCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATT 780  
DB 721 CTTTGTGAAGCCATGCTCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATT 780  
  
QY 781 GGAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCAGTATT 840  
DB 781 GGAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCAGTATT 840  
  
QY 841 AAAATTAAGATCTCACCTTGCTGAACTGATAGGATTAATGATACATGTTTTGCTGT 900  
DB 841 AAAATTAAGATCTCACCTTGCTGAACTGATAGGATTAATGATACATGTTTTGCTGT 900  
  
QY 901 TTGATAACGTGTTAGAGGCCATTCACTGGCACAGACAGTATTACGTCGCTTTACATT 960  
DB 901 TTGATAACGTGTTAGAGGCCATTCACTGGCACAGACAGTATTACGTCGCTTTACATT 960  
  
QY 961 CATAATCCAGACTTTATAGAAATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAA 1020  
DB 961 CATAATCCAGACTTTATAGAAATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAA 1020  
  
QY 1021 ATCTGTGACATTGCAAGGGAATAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTT 1080  
DB 1021 ATCTGTGACATTGCAAGGGAATAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTT 1080  
  
QY 1081 CAGTCAATGACTTATGGATTTAAAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGG 1140  
DB 1081 CAGTCAATGACTTATGGATTTAAAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGG 1140  
  
QY 1141 ATGCTAAAAGATGTGGAGGATGACATGCAAGAAGAGTAAAGAGTACTCGAAGTCGACAA 1200  
DB 1141 ATGCTAAAAGATGTGGAGGATGACATGCAAGAAGAGTAAAGAGTACTCGAAGTCGACAA 1200  
  
QY 1201 GGAGAAGAAAGAGATCCAGAGTTGAACTAGAACACCAACCAATGTTTAGCAGTATTTCAGC 1260  
DB 1201 GGAGAAGAAAGAGATCCAGAGTTGAACTAGAACACCAACCAATGTTTAGCAGTATTTCAGC 1260  
  
QY 1261 AGAGTGAAATTTTACTCGTGTGTTACTGACAGTGTCTATAGCCTTTACTAAGAAAGAGACC 1320  
DB 1261 AGAGTGAAATTTTACTCGTGTGTTACTGACAGTGTCTATAGCCTTTACTAAGAAAGAGACC 1320  
  
QY 1321 AGTCTGTTGCAGAACTCAAAAATTGATGGTTCAAGCAGCAGATCTTTCTGCTCAATT 1380  
DB 1321 AGTCTGTTGCAGAACTCAAAAATTGATGGTTCAAGCAGCAGATCTTTCTGCTCAATT 1380  
  
QY 1381 CATAATTCATTGCATATGGCATCCAGGCCAGAGTATGATACTACAAAAGGAGATCATCCA 1440  
DB 1381 CATAATTCATTGCATATGGCATCCAGGCCAGAGTATGATACTACAAAAGGAGATCATCCA 1440  
  
QY 1441 ATTATGATGGGTTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGA 1500  
DB 1441 ATTATGATGGGTTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGA 1500  
  
QY 1501 TATGCAAAAATAATTAAAAGGGAAGAAATGGTGAACCTATTTTGCAAGATTAAATAGATAGA 1560  
DB 1501 TATGCAAAAATAATTAAAAGGGAAGAAATGGTGAACCTATTTTGCAAGATTAAATAGATAGA 1560  
  
QY 1561 ATAAAAACTGTCTGTGAGGTTGTGAATTTAACAAATTTACATTGTATCTCTGGATTTTTC 1620  
DB 1561 ATAAAAACTGTCTGTGAGGTTGTGAATTTAACAAATTTACATTGTATCTCTGGATTTTTC 1620  
  
QY 1621 TGTGAATTTAGTGAAACAGTCACCATGTGTTCTTTCAAGATCTCTGTATCTCTGGATTTTTC 1680  
DB 1621 TGTGAATTTAGTGAAACAGTCACCATGTGTTCTTTCAAGATCTCTGTATCTCTGGATTTTTC 1680  
  
QY 1681 CTGGTGGATAACAAAAAGGTCTTTGGAACTCACTCATGCAAGACATGGTGAAGATGCA 1740  
DB 1681 CTGGTGGATAACAAAAAGGTCTTTGGAACTCACTCATGCAAGACATGGTGAAGATGCA 1740  
  
QY 1741 CTTCCGTCCTTTTGTGAGATCCTCCGAGTGCTTTTCCCCCAAGTGCTACCTATATAATAATC 1800

1741 CTTGGTCTTTTGTGACATCCTCCGAGTGTCTTCCCCCAAGTGCTACCTATATAATAATC 1800  
1801 ACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTCGGCCATCTCTGTAGTC 1860  
1801 ACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTCGGCCATCTCTGTAGTC 1860  
1861 TTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATCTTGT 1920  
1861 TTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATCTTGT 1920  
1921 AGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTAATAGGACAGAGAAGGT 1980  
1921 AGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTAATAGGACAGAGAAGGT 1980  
1981 TGATGACGCGCTTCACACCATGCTGTGTAACAGGACCCCAAGGCAACATTTGGCCTG 2040  
1981 TGATGACGCGCTTCACACCATGCTGTGTAACAGGACCCCAAGGCAACATTTGGCCTG 2040  
2041 TTATAGGTACCTGGGTCTTTTACCATAACCTTCGCATTTATGATACAGTACCTTCTAAGTGG 2100  
2041 TTATAGGTACCTGGGTCTTTTACCATAACCTTCGCATTTATGATACAGTACCTTCTAAGTGG 2100  
2101 CTTTGAATTTGGAACCTCTACAGTATGCACGAGTACTATATACATATATTGGTATCTCTCTGA 2160  
2101 CTTTGAATTTGGAACCTCTACAGTATGCACGAGTACTATATACATATATTGGTATCTCTCTGA 2160  
2161 ATTCTTTTACGATGTTGATGTCAACATTTGAGTGTGCGGATGCTCTCAATGGCAGA 2220  
2161 ATTCTTTTACGATGTTGATGTCAACATTTGAGTGTGCGGATGCTCTCAATGGCAGA 2220  
2221 GGAAGGATAATGGAAGAGCAGCAGAGAGGCGGTAGTAGTAAAGGCAACAAAGAAAAA 2280  
2221 GGAAGGATAATGGAAGAGCAGCAGAGAGGCGGTAGTAGTAAAGGCAACAAAGAAAAA 2280  
2281 GAAAGTTGCGCCATTTGAGCCGAGAGATCAATGAGCCGAGATATCAGAACATGTGTGC 2340  
2281 GAAAGTTGCGCCATTTGAGCCGAGAGATCAATGAGCCGAGATATCAGAACATGTGTGC 2340  
2341 TGGAATGTTTAAACCATGTTAGTCAATTTGACATGGAGGCAAGTACGTAACCGAAGTT 2400  
2341 TGGAATGTTTAAACCATGTTAGTCAATTTGACATGGAGGCAAGTACGTAACCGAAGTT 2400  
2401 TGAGCTTTGATGTAAGCAAGTTTCGATGTAACACAGGTTTGTCTCATTTCAACAGTGTGAT 2460  
2401 TGAGCTTTGATGTAAGCAAGTTTCGATGTAACACAGGTTTGTCTCATTTCAACAGTGTGAT 2460  
2461 GACCCCGCCGAGTGCATCTTACAGTCAAGGAAATGCTGACCTCAATAAATATAG 2520  
2461 GACCCCGCCGAGTGCATCTTACAGTCAAGGAAATGCTGACCTCAATAAATATAG 2520  
2521 CCCTCCTCCTCAGTCTCCTGAAGTATGTTGCGAGCTAGTAGCACTTTCAACAGGCAAA 2580  
2521 CCCTCCTCCTCAGTCTCCTGAAGTATGTTGCGAGCTAGTAGCACTTTCAACAGGCAAA 2580  
2581 AATGATATTGGAATAATATTCCTAACCCGACCATGAGGTTAATAGAAATTTTAAAGGTTGC 2640  
2581 AATGATATTGGAATAATATTCCTAACCCGACCATGAGGTTAATAGAAATTTTAAAGGTTGC 2640  
2641 CAAACCCAACTTTGTGTTATGAAGTATTTGGCAGGAGGACACAAAAGGAATCTAAAGT 2700  
2641 CAAACCCAACTTTGTGTTATGAAGTATTTGGCAGGAGGACACAAAAGGAATCTAAAGT 2700  
2701 TCCTCCTGAATTTGATTTCTCTGCTCATATAATATTTTCTGTTGTAAGTCTTTGAGA 2760  
2701 TCCTCCTGAATTTGATTTCTCTGCTCATATAATATTTTCTGTTGTAAGTCTTTGAGA 2760  
2761 GAGACTGGGAGGTGGCCATAAAGGGGAGAGTCTTCTTTCAGACCCCACTCTTAGAGGG 2820  
2761 GAGACTGGGAGGTGGCCATAAAGGGGAGAGTCTTCTTTCAGACCCCACTCTTAGAGGG 2820  
2821 CACATCACCAGGCTCCACATCACGGGAAGTGAGATGATTTCTTGGGTAACTCAATTA 2880

2821 CACATCACCAGGCTCCACATCAGGGAAGTGAGATGATGATTTCTTGGGTAACTCAATTA 2880  
2881 TAAGGAATACTTTTAGTTTGAAGCCTTATATGACATGAAATGAACTGCTGTTTAAAG 2940  
2881 TAAGGAATACTTTTAGTTTGAAGCCTTATATGACATGAAATGAACTGCTGTTTAAAG 2940  
2941 TGGTTTATTTATGTTCCATGGAAGAACTGGTCTTATGAAATGCAATGAACTGATAT 3000  
2941 TGGTTTATTTATGTTCCATGGAAGAACTGGTCTTATGAAATGCAATGAACTGATAT 3000  
3001 GGTTTTATTTACAGATTTAATCAAAATCATTTTATGAAATGATTTGAGTGAATAATAGTGT 3060  
3001 GGTTTTATTTACAGATTTAATCAAAATCATTTTATGAAATGATTTGAGTGAATAATAGTGT 3060  
3061 TTATAAAGGTTAATAAATTTCTTGACAAAAA 3096  
3061 TTATAAAGGTTAATAAATTTCTTGACAAAAA 3096

RESULT 2

ABV23940/c  
ID ABV23940 standard; cDNA; 2668 BP.

XX ABV23940;

XX 16-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 23931.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-SEP-2001

XX 20-FEB-2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-183319P.

XX 16-MAR-2000; 2000US-189862P.

XX 25-MAY-2000; 2000US-207454P.

XX 09-JUN-2000; 2000US-211314P.

XX 18-JUL-2000; 2000US-219007P.

XX 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1; Page 4423; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;

did not designate us  
Not Applicable





Db	2068	TGACATTGCAAGGAAAAAGTAAATAAAGCTGCTGTTTTTTGAAGAGGAAGATTTTCAGTC	2009
QY	1086	AATGACTTATGGATTTAAAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGCATGCT	1145
Db	2008	AATGACTTATGGATTTAAAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGCATGCT	1949
QY	1146	AAAAGATGTGGAGGATGACATGCAAAAGAAGAGTAAAGAGTACTCGAAGTCGACAAGGAGA	1205
Db	1948	AAAAGATGTGGAGGATGACATGCAAAAGAAGAGTAAAGAGTACTCGAAGTCGACAAGGAGA	1889
QY	1206	AGAAAGAGATCCAGAAATTGAACCTAGAACACCAACAATGTTTAGCAGTATTCAGCAGAGT	1265
Db	1888	AGAAAGAGATCCAGAAATTGAACCTAGAACACCAACAATGTTTAGCAGTATTCAGCAGAGT	1829
QY	1266	GAAATTTACTCGTGTGTACTGACAGTGCTTTATAGCCTTTACTAAGAAAGAGACCAGTGC	1325
Db	1828	GAAATTTACTCGTGTGTACTGACAGTGCTTTATAGCCTTTACTAAGAAAGAGACCAGTGC	1769
QY	1326	TGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTCTGCCAATTCATAA	1385
Db	1768	TGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTCTGCCAATTCATAA	1709
QY	1386	TTCAATTGCATCATGGCATCCAGGCCCCAGAAATGATACTACAAAAAGGAGATCATCCAATTAT	1445
Db	1708	TTCAATTGCATCATGGCATCCAGGCCCCAGAAATGATACTACAAAAAGGAGATCATCCAATTAT	1649
QY	1446	GATGGGTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGATATGC	1505
Db	1648	GATGGGTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGATATGC	1589
QY	1506	AAAAATAATTAAAAGGGAAGAAATGGTGAACCTATTTTGCAAAGATTAATAGATAGATAAAA	1565
Db	1588	AAAAATAATTAAAAGGGAAGAAATGGTGAACCTATTTTGCAAAGATTAATAGATAGATAAAA	1529
QY	1566	AACGTCTGTGAGGTTGTGAATTTAACAAATTTACATGTATCTCTGGATTTTTTCTGTGA	1625
Db	1528	AACGTCTGTGAGGTTGTGAATTTAACAAATTTACATGTATCTCTGGATTTTTTCTGTGA	1469
QY	1626	ATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTACAAAACCACTTTCCCTGGT	1685
Db	1468	ATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTACAAAACCACTTTCCCTGGT	1409
QY	1686	GGATAACAAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACCTCG	1745
Db	1408	GGATAACAAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACCTCG	1349
QY	1746	GTCTTTTGTGCAGATCCTCCGAGTGCTTTCCCCCAAGTGCTACCTATATAATAATCACAG	1805
Db	1348	GTCTTTTGTGCAG--TCCTCCGGTGCTTTCCCCCAAGTGCTACCTATATAATAATCACAG	1291
QY	1806	GCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTTCGGCCATTCTGTAGCTTTAT	1865
Db	1290	GCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTTCGGCCATTCTGTAGCTTTAT	1231
QY	1866	CAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTTCTTGAGGAA	1925
Db	1230	CAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTTCTTGAGGAA	1171
QY	1926	TTTGCCACCTTGCAGGATGATTTTATGACATTTTATTTAATAGGCAGAGAAGGTTGATG	1985
Db	1170	TTTGCCACCTTGCAGGATG-----AGGCAGAGAAGGTTGATG	1134
QY	1986	CAGCGCTTCACACCATGCTGTGTAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAG	2045
Db	1133	CAGCGCTTCACACCATGCTGTGTAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAG	1074
QY	2046	GTACCTGGCTCCTTTACCATAACTTCGCAATTATGATACAGTACCTTCTAAGTGGCTTTG	2105
Db	1073	GTACCTGGCTCCTTTACCATAACTTCGCAATTATGATACAGTACCTTCTAAGTGGCTTTG	1014
QY	2106	AATTGGAACTCTACAGTATGCACGAG--TACTATTACATATATTTGGTATCTCTCTGAATTC	2164

Db	1013	AATTGGAACTCTACAGTATGCACGAGATACTATTACATATATTGGTATCTCTCTGAATTC	954
QY	2165	CTTTACGCATGGTTGATGTCAACATTGAGTCGTGCCGATGGCTCTCAAATGGCAGAGAA	2224
Db	953	CTTTACGCATGGTTGATGTCAACATTGAGTCGTGCCGATGGCTCTCAAATGGCAGAGAA	894
QY	2225	AGGATAATGGAAAGCAGCAGAGGCGCGGTAGTAAAAAAACAAAGAAAAAAGAAA	2284
Db	893	AGGATAATGGAAAGCAGCAGAGGCGCGGTAGTAAAAAAACAAAGAAAAAAGAAA	834
QY	2285	GTTCGCCCATTTGAGCCGAGAGATCACAAATGAGCCAAAGCATATCAGAACATGTGTCTGGA	2344
Db	833	GTTCGCCCATTTGAGCCGAGAGATCACAAATGAGCCAAAGCATATCAGAACATGTGTCTGGA	774
QY	2345	ATGTTTAAAAACCATGGTAGCATTTGACATGGACGCGCAAAAGTACGTAAACCGAGTTTGAG	2404
Db	773	ATGTTTAAAAACCATGGTAGCATTTGACATGGACGCGCAAAAGTACGTAAACCGAGTTTGAG	714
QY	2405	CTTGATAGTGAACAAGTTCGGTATGAACACACAGGTTTGCTCCATTCAACAGTGTGATGACC	2464
Db	713	CTTGATAGTGAACAAGTTCGGTATGAACACACAGGTTTGCTCCATTCAACAGTGTGATGACC	654
QY	2465	CCGCCGCCAGTGCACTACTTACAGTTCAAGGAATGTCTGACCTCAATAATATAGCCCT	2524
Db	653	CCGCCGCCAGTGCACTACTTACAGTTCAAGGAATGTCTGACCTCAATAATATAGCCCT	594
QY	2525	CCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCAAAAATG	2584
Db	593	CCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCAAAAATG	534
QY	2585	ATATTGAAAAATATTCCTAAACCCGGACCATTAGGTTTAATAGAAATTTTAAAGTTGCCAAA	2644
Db	533	ATATTGAAAAATATTCCTAAACCCGGACCATTAGGTTTAATAGAAATTTTAAAGTTGCCAAA	474
QY	2645	CCCAACTTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCCT	2704
Db	473	CCCAACTTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCCT	414
QY	2705	CCTGAATTTGATTTCTCTGCTCATAAAATATTTTCCCTGTGTGAAACTTGTGAGAGAGA	2764
Db	413	CCTGAATTTGATTTCTCTGCTCATAAAATATTTTCCCTGTGTGAAACTTGTGAGAGAGA	354
QY	2765	CTGGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTTCAGACTTCTTAGAGGGCAC	2824
Db	353	CTGGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTTCAGACCCCAACTCTTAGAGGGCAC	294
QY	2825	TCACCAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAAACAACTCATTATAAG	2884
Db	293	TCACCAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAAACAACTCATTATAAG	234
QY	2885	GAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGTTTTTAAAGTGGT	2944
Db	233	GAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGTTTTTAAAGTGGT	174
QY	2945	TTATTATGTTCCATGGAAAGAACTGGTCTTATTGAATGCATTTGATGAACGTTATATGGTT	3004
Db	173	TTATTATGTTCCATGGAAAGAACTGGTCTTATTGAATGCATTTGATGAACGTTATATGGTT	114
QY	3005	TTATTACAGATTTTAATCACAAATCATTTTTTTTATGAATGATTTGAGTGAAAAATAGTGTAT	3064
Db	113	TTATTACAGATTTTAATCACAAATCATTTTTTTTATGAATGATTTGAGTGAAAAATAGTGTAT	54
QY	3065	AAAGGTTAATAAATTTCTTGACAAAAAATAA 3095	
Db	53	AAAGGTTAATAAATTTCTTGACAAAAAATAA 23	

RESULT 4  
AAL49929  
ID AAL49929 standard; cdna; 2488 BP.  
XX  
AC AAL49929;  
XX

DT 10-DEC-2002 (first entry)  
XX Human molecule for disease detection and treatment coding sequence #3.  
DE  
XX  
KW Human; molecule for disease detection and treatment; MDDT; gene therapy;  
KW cytostatic; antiarteriosclerotic; hepatotropic; anti-HIV; antiallergic;  
KW antiinflammatory; antiasthmatic; cerebroprotective; nootropic;  
KW neuroprotective; antiparkinsonian; cardiant; antiangular; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200270709-A2.  
XX  
PD 12-SEP-2002.  
XX  
PF 08-FEB-2002; 2002WO-US03709.  
XX  
PR 09-FEB-2001; 2001US-268117P.  
PR 15-FEB-2001; 2001US-269618P.  
PR 23-FEB-2001; 2001US-271118P.  
PR 07-MAR-2001; 2001US-274486P.  
PR 09-MAR-2001; 2001US-274436P.  
PR 28-NOV-2001; 2001US-334229P.  
PR 01-FEB-2002; 2002US-353284P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Lal PG, Baughn MR, Yao MG, Walia NK, Elliot VS, Xu Y;  
PI Honchell CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DAM;  
PI Hafalia AJA, Ghandi AR, Thangavelu K, Sanjanwala MM, Tang YT;  
PI Ramkumar J, Griffin JA, Swarnaker A, Azimzai Y, Sapperstein SK;  
PI Burford N, Lee EA, Lu Y, Tran UK, Marquis JP;  
XX  
DR WPI; 2002-713453/77.  
DR P-PSDB; AAO19400.  
XX  
XX New human molecules for disease detection and treatment (MDDT), useful  
PT for diagnosing, treating and preventing diseases or conditions  
PT associated with the aberrant MDDT expression, e.g. cancer, AIDS,  
PT asthma, diabetes, hepatitis -  
XX  
PS Claim 11; Page 159-160; 177pp; English.  
XX  
CC The present invention relates to human proteins and coding sequences of  
CC molecules for disease detection and treatment MDDT. The sequences can be  
CC used in the treatment of diseases associated with the decreased  
CC expression or overexpression of MDDT, such as cell proliferative (cancer,  
CC atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS,  
CC allergies, Addison's disease, asthma), developmental (dwarfism, renal  
CC tubular acidosis), neurological (e.g. stroke, Parkinson's disease,  
CC epilepsy) and cardiovascular (congestive heart failure, myocardial  
CC infarction, angina pectoris) disorders. The present sequence is a  
CC coding sequence of the invention.  
XX  
SQ Sequence 2488 BP; 768 A; 461 C; 567 G; 692 T; 0 other;  
  
Query Match 78.1%; Score 2417.2; DB 24; Length 2488;  
Best Local Similarity 98.5%; Pred. No. 0;  
Matches 2470; Conservative 0; Mismatches 13; Indels 25; Gaps 2;  
  
QY 474 GCGCGGCGGAGCGCGTCTGTTATTTCCGTGGTCCGGACAGTGGTGGCGCGGGT 533  
Db 1 GCGCGGCGGAGCGCGTCTGTTATTTCCGTGGTCCGGACAGTGGTGGCGCGGGT 60  
  
QY 534 GACCACGGGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 593  
Db 61 GACCACGGGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 120  
  
QY 594 ATGGAGCTCAGTATGCCAGAAAAAATGAGAAAAAGCAATACAACTGGGTGGACATTAC 653  
Db 121 ATGGAGCTCAGTATGCCAGAAAAAATGAGAAAAAGCAATACAACTGGGTGGACATTAC 180  
  
QY 654 CCAAGATTTTGAAGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCT 713

Db 181 CCAAGATTTTGAAGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCT 240  
QY 714 ATTTGGTCTTTTGAAGCCATGTCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGG 773  
Db 241 ATTTGGTCTTTTGAAGCCATGTCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGG 300  
QY 774 CATGATTGGAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACCAAGCTATCAAGGATGG 833  
Db 301 CATGATTGGAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACCAAGCTATCAAGGATGG 360  
QY 834 CACTATTAAAAATTAAGATCTCACCTTGCCCTGAACTGATAGGGATTATGGATACATGTTT 893  
Db 361 CACTATTAAAAATTAAGATCTCACCTTGCCCTGAACTGATAGGGATTATGGATACATGTTT 420  
QY 894 TTGCTGTTTGATAACGTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTTACGTGCCT 953  
Db 421 TTGCTGTTTGATAACGTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTTACGTGCCT 480  
QY 954 TTACATTCTAATCCAGACTTTTATAGAAGATCCTTGCTATGAAGGCTTTTGTCTCTGGGAAT 1013  
Db 481 TTACATTCTAATCCAGACTTTTATAGAAGATCCTTGCTATGAAGGCTTTTGTCTCTGGGAAT 540  
QY 1014 CTTGAAAAATCTGTGACATTGCAAGGGGAAAAAGTAAATAAAGCTGCTGTTTTGAAGAGGA 1073  
Db 541 CTTGAAAAATCTGTGACATTGCAAGGGGAAAAAGTAAATAAAGCTGCTGTTTTGAAGAGGA 600  
QY 1074 AGATTTTCAGTCAATGACTTATGGATTAAAAATGGCTAAACAGTGTGACAGATCTTCGAGT 1133  
Db 601 AGATTTTCAGTCAATGACTTATGGATTAAAAATGGCTAAACAGTGTGACAGATCTTCGAGT 660  
QY 1134 TACAGGCATGCTAAAAAGATGTGGAGGATGACATGCAAAAGAGAGTAAAGAGTACTCGAAG 1193  
Db 661 TACAGGCATGCTAAAAAGATGTGGAGGATGACATGCAAAAGAGAGTAAAGAGTACTCGAAG 720  
QY 1194 TCGACAAGGAGAGAAAAAGAGATCCAGAGTTGAACTAGAACACCAACAATGTTTAGCAGT 1253  
Db 721 TCGACAAGGAGAGAAAAAGAGATCCAGAGTTGAACTAGAACACCAACAATGTTTAGCAGT 780  
QY 1254 ATTCAGCAGAGTGAAATTTACTCGTGTGTTACTGACAGTGCTTATAGCCTTACTAAGAA 1313  
Db 781 ATTCAGCAGAGTGAAATTTACTCGTGTGTTACTGACAGTGCTTATAGCCTTACTAAGAA 840  
QY 1314 AGAGACCAGTCTGTTGCAGAAGCTCAAAAATGATGGTTCAAGCAGCAGATCTTCTTTC 1373  
Db 841 AGAGACCAGTCTGTTGCAGAAGCTCAAAAATGATGGTTCAAGCAGCAGATCTTCTTTC 900  
QY 1374 TGCCATTCTAATAATTCATTGTCATCATGGCATCCAGGCCCAGAAATGATACTACAAAAGGAGA 1433  
Db 901 TGCCATTCTAATAATTCATTGTCATCATGGCATCCAGGCCCAGAAATGATACTACAAAAGGAGA 960  
QY 1434 TCATCCAAATATGATGGGTTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTT 1493  
Db 961 TCATCCAAATATGATGGGTTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTT 1020  
QY 1494 CCCTCGATATGCAAAAAATAATTAAGGGAAGAAATGGTGAACCTATTTGCAAGATTAAAT 1553  
Db 1021 CCCTCGATATGCAAAAAATAATTAAGGGAAGAAATGGTGAACCTATTTGCAAGATTAAAT 1080  
QY 1554 AGATAGATAAAAAACTGTCTGTGAGTTGTGAATTTAACAATTTACATTTGATCCTGGA 1613  
Db 1081 AGATAGATAAAAAACTGTCTGTGAGTTGTGAATTTAACAATTTACATTTGATCCTGGA 1140  
QY 1614 TTTTCTGTGAATTTAGTGAACTCACCATGTGTTCTTTCAAGATCTCTGTTACAAAC 1673  
Db 1141 TTTTCTGTGAATTTAGTGAACTCACCATGTGTTCTTTCAAGATCTCTGTTACAAAC 1200  
QY 1674 CACTTTCTGTGGATAACAAAAAGTCTTTTGGAACTCATCTCATGCAAGACATGGTGAA 1733  
Db 1201 CACTTTCTGTGGATAACAAAAAGTCTTTTGGAACTCATCTCATGCAAGACATGGTGAA 1260  
QY 1734 AGATGCACTTCGGTCTTTTGTGACATCCTCCGAGTGCTTTTCCCCCAAGTGCTACCTATAT 1793

Db 1261 AGATGCACCTTCGGTCTTTTGTGTCAGTCCTCCG--GTGCTTTCCCCCRAAGTGCTACCTATAT 1318

Qy 1794 AATAATCACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCAGTGTGTTGGCCATTTC 1853

Db 1319 AATAATCACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCAGTGTGTTGGCCATTTC 1378

Qy 1854 TGTAGTCTTATTCAGATCCATGGACATPACAGGGCTCGACAGAGAGATAAGCTTGGTCAT 1913

Db 1379 TGTAGTCTTATTCAGATCCATGGACATPACAGGGCTCGACAGAGAGATAAGCTTGGTCAT 1438

Qy 1914 ATTCTTGAGGAATTTGCCACCTTGCCAGATGAGTTTATGACATTTTATTTAATAGGCAG 1973

Db 1439 ATTCTTGAGGAATTTGCCACCTTGCCAGATG-----AGGCAG 1475

Qy 1974 AGAAGGTTGATGCGAGCGCTTCCACCATGCTGTTGAAACAGGAACCCCAAAGGCAACATT 2033

Db 1476 AGAAGGTTGATGCGAGCGCTTCCACCATGCTGTTGAAACAGGAACCCCAAAGGCAACATT 1535

Qy 2034 TGGCCTGTTTAGGTACCTGGGTCCCTTTACCATAACTTTCGCATTATGATACAGTACCTTC 2093

Db 1536 TGGCCTGTTTAGGTACCTGGGTCCCTTTACCATAACTTTCGCATTATGATACAGTACCTTC 1595

Qy 2094 TAAGTGGCTTTGAATTGGAATCTACAGTATGCACGAGTACTATTACATATATTGGTATC 2153

Db 1596 TAAGTGGCTTTGAATTGGAATCTACAGTATGCACGAGTACTATTACATATATTGGTATC 1655

Qy 2154 TCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTGAGTCGTGCCGATGGCTCTCAAA 2213

Db 1656 TCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTGAGTCGTGCCGATGGCTCTCAAA 1715

Qy 2214 TGGCAGAGGAAGGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAACAAAGA 2273

Db 1716 TGGCAGAGGAAGGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAACAAAGA 1775

Qy 2274 AAAAAAGAAAGTTGCCCATTTGAGCCGAGAGATCAAAATGAGCCACAGCATATCAGAACA 2333

Db 1776 AAAAAAGAAAGTTGCCCATTTGAGCCGAGAGATCAAAATGAGCCACAGCATATCAGAACA 1835

Qy 2334 TGTGTGCTGGAATGTTTAAACCATGGTAGCATTTGACATGGACGGCAAGTACGTAAC 2393

Db 1836 TGTGTGCTGGAATGTTTAAACCATGGTAGCATTTGACATGGACGGCAAGTACGTAAC 1895

Qy 2394 CGAAGTTTGAGCTTGATAGTGAACAAGTTTCGGTATGAACACAGGTTTGTCTCCATTCAACA 2453

Db 1896 CGAAGTTTGAGCTTGATAGTGAACAAGTTTCGGTATGAACACAGGTTTGTCTCCATTCAACA 1955

Qy 2454 GTGTGATGACCCCGCCGCGAGTGCACCTACTTACAGTTCAAGGAAATGCTGACCTCAATA 2513

Db 1956 GTGTGATGACCCCGCCGCGAGTGCACCTACTTACAGTTCAAGGAAATGCTGACCTCAATA 2015

Qy 2514 AATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACTTTCAAC 2573

Db 2016 AATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACTTTCAAC 2075

Qy 2574 AGGCAAAAATGATATTGGAAAAATATTCCTAACCCGGACCATGAGGTTAATAGAAATTTAA 2633

Db 2076 AGGCAAAAATGATATTGGAAAAATATTCCTAACCCGGACCATGAGGTTAATAGAAATTTAA 2135

Qy 2634 AGGTTGCCAAACCCCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAAT 2693

Db 2136 AGGTTGCCAAACCCCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAAT 2195

Qy 2694 CTAAGTTCCTCTGAATTTGATTTCTCTGCTCATAAATATTTTCTGTGTGAAACTTG 2753

Db 2196 CTAAGTTCCTCTGAATTTGATTTCTCTGCTCATAAATATTTTCTGTGTGAAACTTG 2255

Qy 2754 TTTGAGAGAGACTGGGAGGTGGCCATAAAGGGCAGAGTCTTCTTTTCAGACCCCACTCT 2813

Db 2256 TTTGAGAGAGACTGGGAGGTGGCCATAAAGGGCAGAGTCTTCTTTTCAGACCCCACTCT 2315

Qy 2814 TAGAGGGCACATCACCAGGCTCCACATCAGCGGAAGTGAGATGGATTCTTGGGTAACAA 2873

Db 2316 TAGAGGGCACATCACCAGGCTCCACATCAGCGGAAGTGAGATGGATTCTTGGGTAACAA 2375

Qy 2874 CTCATTATAAGGAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGT 2933

Db 2376 CTCATTATAAGGAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGT 2435

Qy 2934 TTTAAAGTGGTTTATTATGTTCCATGGAAGAAACTGGTCTTATTGAAT 2981

Db 2436 TTTAAAGTGGTTTATTATGTTCCATGTAAGACACTGGGTTCCATTAAAT 2483

RESULT 5

ABS51309

ID ABS51309 standard; cDNA; 719 BP.

XX ABS51309;

AC ABS51309;

XX ABS51309;

DT 21-OCT-2002 (first entry)

XX cDNA encoding human secretory protein #7.

DE Human; secretory polypeptide; SPTM; actinic keratosis; arteriosclerosis; bursitis; cirrhosis; hepatitis; hepatitis; polycythaemia vera; anaemia; psoriasis;

XX KW primary thrombocytopenia; cancer; adenocarcinoma; leukaemia; myeloma;

KW sarcoma; immune system disorder; acquired immunodeficiency syndrome;

KW AIDS; allergy; asthma; Crohn's disease; diabetes mellitus; gout;

KW glomerulonephritis; Goodpasture's syndrome; thyroiditis; pancreatitis;

KW hepatitis; multiple sclerosis; osteoporosis; Reiter's syndrome;

KW rheumatoid arthritis; neurological disorder; epilepsy; stroke; dementia;

KW Alzheimer's disease; Pick's disease; Huntington's disease; mood; anxiety;

KW Parkinson's disease; central nervous system disorder; mental disorder;

KW schizophrenic disorder; amnesia; Tourette's disorder; transgenic animal;

KW gene therapy; gene; ss.

XX Homo sapiens.

OS Homo sapiens.

XX WO200257304-A2.

PN 25-JUL-2002.

XX 15-JAN-2002; 2002WO-US01340.

PD 16-JAN-2001; 2001US-261864P.

PR 16-JAN-2001; 2001US-261865P.

PR 16-JAN-2001; 2001US-261979P.

PR 16-JAN-2001; 2001US-261981P.

PR 17-JAN-2001; 2001US-262164P.

PR 17-JAN-2001; 2001US-262208P.

PR 17-JAN-2001; 2001US-263131P.

PR 19-JAN-2001; 2001US-262599P.

PR 19-JAN-2001; 2001US-262760P.

PR 19-JAN-2001; 2001US-263063P.

PR 19-JAN-2001; 2001US-263066P.

PR 19-JAN-2001; 2001US-263069P.

PR 19-JAN-2001; 2001US-263070P.

PR 19-JAN-2001; 2001US-263074P.

PR 19-JAN-2001; 2001US-263076P.

PR 19-JAN-2001; 2001US-263077P.

PR 19-JAN-2001; 2001US-263329P.

XX (INCY-) INCYTE GENOMICS INC.

PA Panzer SR, Lincoln SE, Altus CM, Dufour GE, Hillman JL, Jones AL;

XX Dam TC, Liu TF, Harris B, Flores V, Daffo A, Marwaha R, Chen AJ;

XX Chang SC, Gerstin EH, Peralta CH, David MH, Lewis SA;

DR WPI; 2002-590716/63.

DR P-PSDB; ABG69816.

XX New purified secretory polypeptides and polynucleotides, useful in the

PT diagnosis, study, prevention or treatment of diseases associated with

PT decreased expression of functional secretory molecules, e.g. AIDS,

PT cancer or allergies -

XX

PS Claim 1; Page 247; 340pp; English.

XX The invention describes an isolated polynucleotide a naturally occurring

CC polynucleotide sequence at least 90 % identical to it, a polynucleotide

CC complementary to it or an RNA equivalent of it. The purified secretory

CC polypeptides (SPTM) and polynucleotides are useful in the diagnosis,

CC study, prevention or treatment of diseases associated with decreased

CC expression of functional SPTM, e.g. actinic keratosis, arteriosclerosis,

CC bursitis, cirrhosis, hepatitis, polycythaemia vera, primary

CC thrombocytopenia, anaemia, psoriasis, cancers including adenocarcinoma,

CC leukaemia, myeloma or sarcoma, immune system disorder such as acquired

CC immunodeficiency syndrome (AIDS), allergies, asthma, Crohn's disease,

CC diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout,

CC Hashimoto's thyroiditis, hepatitis, multiple sclerosis, osteoporosis,

CC pancreatitis, Reiter's syndrome, autoimmune thyroiditis or rheumatoid

CC arthritis, neurological disorders such as epilepsy, stroke, Alzheimer's

CC disease, Pick's disease, Huntington's disease, dementia, Parkinson's

CC disease, other developmental disorder of the central nervous system,

CC mental disorder including mood, anxiety or schizophrenic disorder,

CC amnesia or Tourette's disorder. The polynucleotides may be used in

CC hybridisation and amplification technologies, e.g. in assessing gene

CC expression patterns, to develop a transcript image for a particular cell

CC or tissue, or to create transgenic animals to model human disease. This

CC sequence encodes a human secretory protein isolated in the invention.

XX

SQ Sequence 719 BP; 223 A; 113 C; 188 G; 195 T; 0 other;

Query Match 23.2%; Score 717.4; DB 24; Length 719;

Best Local Similarity 99.9%; Pred. No. 7e-169;

Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 474 GGCGGCGCGGAGGCGGCGTCTTATTTCCGTGGTCCGACAGTCCGTGGCGCGGGT 533

DB 1 GGCGGCGCGGAGGCGGCGTCTTATTTCCGTGGTCCGACAGTCCGTGGCGCGGGT 60

QY 534 GACCACGGGAGAGTAGGCATATGTTATGAAAGCTTCTGTAGATGATGACGATTGAGG 593

DB 61 GACCACGGGAGAGTAGGCATATGTTATGAAAGCTTCTGTAGATGATGACGATTGAGG 120

QY 594 ATGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTAC 653

DB 121 ATGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTAC 180

QY 654 CCAAGATTTTGAAGAAGCTTGTGCGAATTAAGTTGGGAACTACTTTCATGATAAGCT 713

DB 181 CCAAGATTTTGAAGAAGCTTGTGCGAATTAAGTTGGGAACTACTTTCATGATAAGCT 240

QY 714 ATTTGGTCTTTTGAAGCCATGCTGCTATTGAAATGATGATGCCAAGATGGATGCTGG 773

DB 241 ATTTGGTCTTTTGAAGCCATGCTGCTATTGAAATGATGATGCCAAGATGGATGCTGG 300

QY 774 CATGATTGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGATGG 833

DB 301 CATGATTGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGATGG 360

QY 834 CACTATTAAATTAAGATCTCACCTTGCTGAACTGATAGGATTTATGGATACATGTTT 893

DB 361 CACTATTAAATTAAGATCTCACCTTGCTGAACTGATAGGATTTATGGATACATGTTT 420

QY 894 TTGCTGTTGATAACGTTAGAGGCGCATTCACCTGGCCACACAGATTTTACGTCCT 953

DB 421 TTGCTGTTGATAACGTTAGAGGCGCATTCACCTGGCCACACAGATTTTACGTCCT 480

QY 954 TTACATTCAATCCAGACTTTATAGAGATCCTGCTATGAGGCTTTTGCTCTGGGAAT 1013

DB 481 TTACATTCAATCCAGACTTTATAGAGATCCTGCTATGAGGCTTTTGCTCTGGGAAT 540

QY 1014 CTTGAAATCTGTGACATTCGCAAGGAAAAAGTAAATAAGTCTGTTTTTGAAGAGGA 1073

DB 541 CTTGAAATCTGTGACATTCGCAAGGAAAAAGTAAATAAGTCTGTTTTTGAAGAGGA 600

QY 1074 AGATTTTCAGTCAATGACTTATGGATTTAAATGGCTAACAGTGTGACAGATCTTCGAGT 1133

Db 601 AGATTTTCAGTCAATGACTTATGATTTAAATGGCTACAGTGTGACAGATCTTCGAGT 660

QY 1134 TACAGGCATGCTAAAGAGATGTGGAGGATGACATGCAAAAGAGAGTAAAGAGTACTCGAA 1192

Db 661 TACAGGCATGCTAAAGAGATGTGGGGATGACATGCAAAAGAGAGTAAAGAGTACTCGAA 719

RESULT 6

ABX50281

ID ABX50281 standard; cDNA; 398 BP.

XX

AC ABX50281;

XX

DT 25-FEB-2003 (first entry)

XX

DE Bovine EST associated with lactation/muscle/fat deposition #210.

XX

KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;

KW muscle deposition; fat deposition; genome mapping; gene identification;

KW gene analysis; cattle breeding.

XX

OS Bos Taurus.

XX

PN US2002137160-A1.

XX

PD 26-SEP-2002.

XX

PF 26-OCT-2001; 2001US-0983965.

XX

PR 17-DEC-1998; 98US-113678P.

PR 15-DEC-1999; 99US-0465231.

XX

PA (BYAT/) BYATT J C.

PA (MATH/) MATHIALAGAN N.

PA (TAON/) TAO N.

PA (WARR/) WARREN W C.

XX

PI Byatt JC, Mathialagan N, Tao N, Warren WC;

XX

PI WPI; 2003-102386/09.

XX

PT Purified nucleic acid molecules, useful for genome mapping, gene

PT identification and analysis, cattle breeding or preparation of

PT constructs for cattle gene expression and genetically improved cattle -

XX

PS Claim 2; SEQ ID No 210; 38pp; English.

XX

CC The invention relates to a purified nucleic acid molecule associated with

CC lactation or muscle and fat deposition (designated LMFD), derived

CC from cattle, and the LMFD nucleic acid can specifically hybridise to a

CC second nucleic acid molecule comprising any of 5912 nucleotide

CC sequences, appearing as ABX50072-ABX55983, or complements of them.

CC Also included are; (1) a transformed cell having a nucleic acid

CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-

CC translated sequence that functions in the cell to cause termination of

CC transcription and addition of polyadenylated ribonucleotides to a 3' end

CC of the mRNA molecule; and (2) determining a level or pattern of a

CC molecule in a bovine cell or tissue comprising: (a) incubating a marker

CC nucleic acid (comprising any of the 5912 nucleic acid sequences or its

CC complement or fragment) with a complementary nucleic acid molecule

CC obtained from the bovine cell or tissue, where hybridisation between the

CC marker nucleic acid and the complementary nucleic acid permits the

CC detection of the molecule; and (b) detecting the level or pattern of the

CC complementary nucleic acid, where the detection of the complementary

CC nucleic acid is predictive of the level or pattern of the molecule.

CC The LMFD nucleic acid is used for determining a level or pattern

CC of a molecule in a bovine cell or tissue. It is useful for genome

CC mapping, gene identification and analysis, cattle breeding, preparation

CC of constructs for use in cattle gene expression, or for genetically

CC improving cattle. The present sequence is one of the 5912 bovine

CC LMFD EST (expressed sequence tag) nucleic acids.

CC Note: The present sequence was not shown in the specification but

CC was obtained in electronic format from the USPTO web site:







PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
XX WPI; 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer --  
XX  
PS Claim 1; Page 8947; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 439 BP; 113 A; 104 C; 110 G; 112 T; 0 other;  
  
Query Match 9.3%; Score 288.8; DB 23; Length 439;  
Best Local Similarity 91.4%; Pred. No. 6.8e-62;  
Matches 338; Conservative 0; Mismatches 7; Indels 25; Gaps 2;  
  
QY 1687 GATAACAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACTTCGG 1746  
Db |||||  
QY 439 GATAACAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACTTCGG 380  
Db |||||  
QY 1747 TCTTTTGTGAGATCCTCCGAGTGTCTTCCCCCAAGTGCTACCTATATATAATCACCAGG 1806  
Db |||||  
QY 379 TCTTTTGTGAG--TCTCCGGTGTCTTCCCCCAAGTGCTACCTATATATAATCACCAGG 322  
Db |||||  
QY 1807 CTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTCCGCCATTCTGTAGTCTTATTC 1866  
Db |||||  
QY 321 CTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTCCGCCATTCTGTAGTCTTATTC 262  
Db |||||  
QY 1867 AGATCCATGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAATTTCTTGAGGAAT 1926  
Db |||||  
QY 261 AGATCCATGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAATTTCTTGAGGAAT 202  
Db |||||  
QY 1927 TTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTAATAGGACAGAAAGGTTGATGC 1986  
Db |||||  
QY 201 TTGCCACCTTGCAGGATG-----AGGCAGAGAAAGGTTGATGC 165  
Db |||||  
QY 1987 AGCGTTTCACACCATGCTGTTGAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGG 2046  
Db |||||  
QY 164 AGCGTTTCACACCATGCTGTTGAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGG 105  
Db |||||  
QY 2047 TACCTGGGTC 2056  
Db |||||  
QY 104 TACCTCGGCC 95  
Db |||||  
  
RESULT 12  
ABQ51660/c  
ID ABQ51660 standard; DNA; 799 BP.  
XX  
AC ABQ51660;

XX 12-JUL-2002 (first entry)  
XX  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 38251.  
DE  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200218632-A2.  
PN  
XX  
XX 07-MAR-2002.  
PD  
XX  
XX 01-SEP-2001; 2001WO-BP10074.  
PF  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
PR  
XX 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
XX WPI; 2002-371829/40.  
DR  
XX  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -  
XX  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
PS  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
SQ Sequence 799 BP; 122 A; 95 C; 262 G; 320 T; 0 other;  
  
Query Match 8.9%; Score 274.2; DB 24; Length 799;  
Best Local Similarity 73.9%; Pred. No. 4.1e-58;  
Matches 348; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
  
QY 86 CTGCTCGTTCGTCTCGCAACCACTAAGGTCTACGCAACCTCCACGGTTTCTTCCGCC 145  
Db |||||  
QY 799 CTACTCGTTCGTACTCGCAACCACTAATAATCTACGCAACCTCCACGATTTCTTCCGCC 740  
Db |||||  
QY 146 TTCCGCTCACCTTTCTAAGAAATTTCCAGAGGGCAGCGCAGCGGGCGGCTCTGAGAC 205  
Db |||||  
QY 739 TTCCGCTCACCTTTCTAATAAATTTCCAAAAACAAACGCAACGAACTCTAAAC 680  
Db |||||  
QY 206 TCCGGGCTCCGCCTCTTTCCGGAACCGCCCTACTACCCAGGACTCCGACAGGGTGAAA 265  
Db |||||  
QY 679 TCCGAACCTCCGCCTCTTTCCGAAAAACCGCCCTACTACCCAAAACTCCGACAAATAATAAAA 620  
Db |||||  
QY 266 AAAGATAACTTCCGGTCTCGCGATCGTCTCTAATCTCGCGAGAAAGAGCGCGGCCCA 325  
Db |||||

Db 619 AAAAATAACTTCGATCGCGATCGTCTCTAATCTCGGAAAAAAGACCGCCA 560  
QY 326 TCGGCCGAACGGAGCGGTGGCGAGGAGGGGGTGTGGCCGGGAGCGCGAAGTCCCGG 385  
Db 559 TCGACCGAACGAAACGATAACGAAAAAATAATACCGAAAAACGCGAATCCCGA 500  
QY 386 GAGTAAGGAGAGGGCGGGGTGCGCGCTCCCGGCATACGCATGCGTGCAAGTCCCG 445  
Db 499 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 440  
QY 446 GTCGGGCTGGGCTGAGAGGGGAGGGGGCGGGCGGCGGAGCGCGGCTGTTTCCGT 505  
Db 439 ATCGAACTAACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 430  
QY 506 GGTCCGGACAGTGGCGGGCGGGGTGACCGACCGGAGAGTAGGCATAA 556  
Db 379 AATCCGAACAATACGTAAACGACGCGAATAACCGAAAAAATAAATAA 329

RESULT 13  
ABQ51661  
ID ABQ51661 standard; DNA; 799 BP.  
XX  
AC ABQ51661;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 38252.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI; 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX

This invention describes a novel method for determining the degree of  
methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
genomic sample of DNA. The sample is treated chemically to convert  
cytosine (C) but not methylated C, to uracil, then part of the genomic  
DNA that contains the target C is amplified to form a labeled amplicon.  
The amplicon is hybridised to two classes, each with at least one  
member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
and the degree of hybridisation to both classes is determined from the  
label on the amplicon. From the ratio of labels hybridised to the two  
classes of oligomers, the degree of methylation is calculated. The method  
is used: (i) for diagnosis and/or prognosis of side effects of  
therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
of the central nervous, cardiovascular, gastrointestinal and respiratory  
systems etc., particularly by detecting mutations or single nucleotide  
polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
SQ Sequence 799 BP; 320 A; 262 C; 95 G; 122 T; 0 other;  
Query Match 8.9%; Score 274.2; DB 24; Length 799;  
Best Local Similarity 73.9%; Pred. No. 4.1e-58;  
Matches 348; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
QY 86 CTGCTCGTTCGTGCTCGCAACCACTAAGGCTCTACGCAAACTCCACGGTTTCTTCCGCC 145  
Db 1 CTACTCGTTCGTACTCGCAACCACTAAATCTACGCAAACTCCACGATTTCTTCCGCC 60  
QY 146 TTCGCGTCACCTTTCTAAGAAATTTCCAGAGGCGAGCGCAGACGGGGCGGCTCTGAGAC 205  
Db 61 TTCGCGTCACCTTTCTAATAAATTTCCCAAAAAACAACGCAACGAACTCTAAAC 120  
QY 206 TCCGGGCTCCGCCTCTTTCCGGGAACCGCCCACTACCCAGGACTCCGACAGAGGGTGAAA 265  
Db 121 TCCGAACTCCGCCTCTTTCCGAAACCGCCCACTACCCAAACTCCGACAAAAATAAAA 180  
QY 266 AAAGATAACTTCCGGTCTCGCGATCGTCTCTAATCTCGGAGAGAGAGAGCGGCGCCCA 325  
Db 181 AAAATAAATACTTCGGATCTCGCGATCGTCTCTAATCTCGGAAAAAATAAATAAATAA 240  
QY 326 TCGGCCGAACGGAGCGGTGGCGAGGGAGGGGTGTGTGGCCGGGAGCGCGAAGTCCCGCG 385  
Db 241 TCGACCGAACGAAAAACGATAAACGAAAAAATAAATAAATAAATAAATAAATAAATAA 300  
QY 386 GAGTAAGGAGAGGGGGCGGGGTTCGCGCTCCCGGCATACGCATGCGTGACGCTGCCG 445  
Db 301 AATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 360  
QY 446 GTCGGGCTGGGCTGAGAGGGGAGGGGGCGGGCGGCGCGCGCGCGCGCGCGCGCGT 505  
Db 361 ATCGAACTAACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 420  
QY 506 GGTCCGGACAGTGGTGGCGGCGGGGTGACCGACCGGAGAGTAGGCATAA 556  
Db 421 AATCCGAACAATACGTAAACGACGCGAATAACCGAAAAAATAAATAAATAAATAA 471

RESULT 14  
ABV06162/c  
ID ABV06162 standard; cDNA; 303 BP.  
XX  
AC ABV06162;  
XX  
DT 13-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 6153.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US05171.  
XX  
PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX



QY 1119 GACAGATC---TTCGAGTTACAGGCATGCTAAAAAGATGTGAGGATGACATGCAAGAAG 1175  
Db 830 GGCACAGCCAGCCACTGTTTATAGCTCGCTAAAGGATGTGAGGATGAGCTCATACG--- 886  
QY 1176 AGTAAAGAGTACTCGAAGTCGACAAGGAGAAGAAAGAGATCCAGAAGTTGAACTAGAAC 1235  
Db 887 -----CAAGTGCAAAAAGCTGACTTCCACGGAGGACTGGATGG 924  
QY 1236 CCAACAATGTTTAGCAGTATTTACGACAGAGTGAAATTTACTCGTGTGTTACTGACAGTGCT 1295  
Db 925 CCGTGGTTCACGGCTGCGTT--TCATGCGCCATCTTTTCAGGTTGATCTACCACGTGGA 982  
QY 1296 TATAGCCTTTACTAAGAAAGAGACCAGTGTCTTGTGCAGAGCTCAAAAATTTGATGTTCA 1355  
Db 983 GCAATGGCCAGCAATGATACCGTCGATGACAAGGTAGACATATACAAAATCCTGCTGT 1042  
QY 1356 AGCAGCAGATCTTCTTCTGCCAFTCATAAATTCATTCATGTCATGGCATCCAGGCCAGAA 1415  
Db 1043 GGCCTCCGAGATGCTTCCAGGAATAAGGAACACTTTGGATCGGGGCACACAACCGAGAA 1102  
QY 1416 TGATACTACAAAAGGAGATCATCCAATTATGATGGGTTTTGAACCCCTTGTGAACCGAG 1475  
Db 1103 GGGGTCT-----GACGCTCCCAATCCCATGGGTTTCTTCGCGCATCCACGACCG 1153  
QY 1476 GCTACTTCCACCTACCTTCCCTCGATATGCAAAAATAATTAAGGGAAGAAATGGTGAA 1535  
Db 1154 CAGCCAACCGCCCGGCTTTCGCGTAGCATTAAGATCAGGGATCGTCCATCCAGTTATCA 1213  
QY 1536 CTATTTTGCAAGATTAAATAGATAGAATAAAAACTGTCTGTGAGGTTGTGAATTTAACAA 1595  
Db 1214 GTTCTAGAGGAATGATTTCCGCGTTCAATACGCCCTGCAAAAGTCACCAAGTACAAGGA 1273  
QY 1596 TTTACATTGTATCCTGGATTTTCTGTGAATTTAGTGAACAGTC---ACCATGTGTTCT 1652  
Db 1274 TTAATAATCGGCGCTGAACCTCTTTATCGAGTACAGCAAAAAGTCGGGCCAGTGCATCCT 1333  
QY 1653 TTCAAGATCTCTGTTACAAACCACTTTCCTGGTGGATAACAAAAGGTCTTTGGAACTCA 1712  
Db 1334 GTCCAGAAGCGTGTGTCAAAACCGCTGTTCAGCGCCAACTGCGTATGGCGCACGGAAAGCT 1393  
QY 1713 TCTCATGCAAGACATGGTGAAGATGCATTCGGTCTTTTGTGATCCTCCGAGTGCTT 1772  
Db 1394 TCCCATGAAGCAGTTCCCTGGCCACTCGGTTCAAGTCTTCAACTCGCCACCCGTTATTGAA 1453  
QY 1773 TCCCCCAAGTGCTACCTATATAATAATCACCAAGCTTAAGCACTGTATCGACTCCTTTGTT 1832  
Db 1454 T--GCCAAGCATCCGGTGGCTGCCGATCCCAAGGTGACGACAGCACCTGGAGAACTTCTC 1511  
QY 1833 ACTCACTGTGTT-----CGGCCATTCTGTAGTCTTATTCAGATCCATGGACATAACAGG 1886  
Db 1512 CGTTACTGCATTAAACATGAACACCTTCAACGAGTTCTATCGCATCTGGGCTTCAATCGC 1571  
QY 1887 GCTCGACAGAGAGATAAGCTTGGTCATATTTCTTGAGGAATTTGCCACCTTGACGATGAG 1946  
Db 1572 GCCAGACAGCGGGATAAGCTGGCGGCTCTGATAGAGAACTTCGACACCATCCAAAGTGGAT 1631  
QY 1947 TTTATGACATTTTATTTAATAGGCAGAGAAGTTGATGCGAGCGCTTCACACCATGCTGT 2006  
Db 1632 GCAGCGGATTTGGACTCTATGATGAACCAACTGGCCAAACGAACGAGCCATGGAGGGCAAC 1691  
QY 2007 TGAAACAGGAACCCCAAG-GCAACATTTGGCCTGTTTAGGTACCTGGGTCTTTTACCAT 2065  
Db 1692 GAGCCCATGGCCACCGCTCTGAAGCACAGCACCCCACTTCTCCACATGGGTGCTGTACAAC 1751  
QY 2066 AACCTTCGCATTTATGATACAGTACCTTCTAAGTGGCTTTGAATTTGGAACCTACAGTATG 2125  
Db 1752 TGTTTCCGTGCGATGCTGATCTTCTGTATGTCCGGCTTTGAGCTGGAACATATGCTGTG 1811  
QY 2126 CACGAGTACTATTACATATATTTGATCTCTCTGAATTCCTTTACGCATGGTTGATGTCA 2185  
Db 1812 CATGAGTTCCCTCTACATCTACTGGTATCCATACGAGTTCCTCATTTGGCTTCCTGCTCC 1871  
QY 2186 ACATTGAGTCGTGCCGATGGCTCTCAAAATGGCAGAGGAAGGATAATGGAAAGCAGCAG 2245

Db 1872 GCGCTGACGCGCACGGAGAAACATCTCTCTGGCCCAGGAGGAGTACGCCGACCAAAAGT 1931  
QY 2246 AAAGGCCGTAGTAG-----TAAAAAACAAAG 2272  
Db 1932 AAGACCCAGTCGGCGCGGCGAGCGGCTAAAAACCGAAGGCGGCCAAGCCGAAAAAG 1991  
QY 2273 AAAAAAAGAAAGTTTCGCCCATTTGAGCCGAGAGATCAACATGAGCCAAAGCATATCAGAAC 2332  
Db 1992 AACAAAAAGACTCAAAAGGCGGTACCGCGCCGAGATCGTCTTCTACACGCGTACTCAGC 2051  
QY 2333 ATGTGTGCTGGAATGTTTAAACCATGGTAGCATTTGACATGGACGGCAAGTACGTAAA 2392  
Db 2052 CTCTGCGCGGAATGTATAAAGCAATGGCGCCCTGACCAAAAGACGGACGCGTGCAGCTG 2111  
QY 2393 CCGAAGTTTGAGCTTGATAGTGAACAAAGTTCGGTATGAACACAGGTTTGTCTCCATTCAAC 2452  
Db 2112 CCGCTGTCCAAGTTCGACAACAGGAGATCCGCTACAAACGACGCTTCTTCCCTTCGCC 2171  
QY 2453 AGTGTGATGACCCCGCCGCGAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAAT 2512  
Db 2172 ACACTCACCAAGTCTCCGCCAGTTTCGTACGACAGTTTCAAAAACGTCGGGAGCACATG 2231  
QY 2513 AAATATAGCCCTCCTCCTCAGTCTCCTGAACGTGTATGTGCGAGCTAGTAAGCACCTTTCAA 2572  
Db 2232 A-----TGCACATAGCGTCGAGGATCTATATACCTATGCGGCCAAGCACTTTGAC 2282  
QY 2573 CAGGCAAAAATGATATTTGMAAAATATTCCTAACCCCGACCATGAGGTTAATAGAATTTTA 2632  
Db 2283 CAGGCCCGCAATGTTCTCGAAAGCATTCAAAACCCAGATCAGGAGATGCTGGACCTCCTG 2342  
QY 2633 AAGTTGCCAAACCCAACTTTTGTGTTATGAAGTTATGGCAGGAGGACACAAAAAGGAA 2692  
Db 2343 CAAATCGCACGGACTAACTTTGTGGTGATGAATGTGCTGGCTCGCGGTCAACCAGAAGGAA 2402  
QY 2693 TCTAAAGTTCTCCTGAAATTTGATTTCTCTGCTCATAAAATATTTTCTGTTGTGAAACT 2751  
Db 2403 GTCAAGCGGCAGCCAGGTTTCGACTTCTCCAAACACACAGCTACTTCCCCATCATCAAGCT 2461

Search completed: November 24, 2003, 22:01:07  
Job time : 790 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2003, 21:46:30 ; Search time 11138 Seconds  
(without alignments)  
11371.539 Million cell updates/sec

Title: US-10-001-857-42  
Perfect score: 3096  
Sequence: 1 ttctctcacgaactccagg.....atttcttgacaaaaaaaaa 3096

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3096	100.0	3096	6	AX523454	AX523454 Sequence
2	2417.2	78.1	2488	6	AX537452	AX537452 Sequence
3	1838.4	59.4	2552	10	AY102701	AY102701 Mus muscu
4	1799.2	58.1	2543	10	AF272892	AF272892 Rattus no
5	1766.6	57.1	1839	9	AK025266	AK025266 Homo sapi
6	1264.8	40.9	2467	5	BC053286	BC053286 Danio rer
7	988	31.9	997	9	AK026296	AK026296 Homo sapi
8	885.4	28.6	1796	9	AK056059	AK056059 Homo sapi
9	717.4	23.2	719	6	AX664135	AX664135 Sequence
10	664.2	21.5	147876	2	AC102010	AC102010 Mus muscu
11	551.4	17.8	181864	9	AL353743	AL353743 Human DNA
12	398.4	12.9	160115	9	AL161447	AL161447 Human DNA
13	321.4	10.4	446	9	HSA334821	AJ334821 Homo sapi
14	299	9.7	743	9	HSA334413	AJ334413 Homo sapi
15	236.4	7.6	642	9	HSA342328	AJ342328 Homo sapi
16	236.4	7.6	742	9	HSA331811	AJ331811 Homo sapi
17	233.2	7.5	640	9	HSA342329	AJ342329 Homo sapi
18	233.2	7.5	652	9	HSA342327	AJ342327 Homo sapi
19	217	7.0	2537	3	AY060948	AY060948 Drosophil
20	207.6	6.7	2565	3	AY089452	AY089452 Drosophil
21	207	6.7	320902	2	AL161456	AL161456 Homo sapi
22	182	5.9	489	11	G63197	G63197 SHGC-140830
23	177.8	5.7	34935	9	AL161453	AL161453 Human DNA
24	177.6	5.7	111554	2	AC143812	AC143812 Macaca mu
25	155.8	5.0	113836	10	AL669837	AL669837 Mouse DNA
26	153.2	4.9	253611	2	AC111473	AC111473 Rattus no
27	153.2	4.9	271861	2	AC108541	AC108541 Rattus no
28	153.2	4.9	281804	2	AC134869	AC134869 Mus muscu
29	140.4	4.5	151	6	AX523453	AX523453 Sequence
30	134.4	4.3	204131	2	AC128898	AC128898 Rattus no
31	130.4	4.2	3497	10	AY102702	AY102702 Mus muscu
32	130.4	4.2	269131	2	AC135469	AC135469 Mus muscu
33	90.6	2.9	16016	2	AC020508	AC020508 Drosophil
34	90.6	2.9	60135	2	AC091200	AC091200 Drosophil
35	90.6	2.9	164443	3	AC012167	AC012167 Drosophil
36	90.6	2.9	298827	3	AE003463	AE003463 Drosophil
37	87.8	2.8	815	6	AX523523	AX523523 Sequence
38	69.4	2.2	7218	6	I66494	I66494 Sequence 14
39	66.6	2.2	2000	6	AX655393	AX655393 Sequence
40	66.6	2.2	110737	2	AC011105	AC011105 Homo sapi
41	66.2	2.1	159980	2	AP005743	AP005743 Oryza sat
42	65.6	2.1	68328	2	AC137705	AC137705 Homo sapi
43	64.6	2.1	832	11	BV026280	BV026280 S212P6647
44	64.4	2.1	69373	2	AC101456	AC101456 Mus muscu
45	64	2.1	67769	2	AC068336	AC068336 Homo sapi

ALIGNMENTS

RESULT 1  
AX523454

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AX523454  
Sequence 42 from Patent WO02064788.  
AX523454  
AX523454.1 GI:24412350

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1

Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.

Compositions and methods relating to lung specific genes and proteins

3096 bp DNA linear PAT 24-OCT-2002



QY	1981	1981	TGATGCAGCGCTTACACCATGCTGTTGAAACAGGAACCCCAAAGGCAACATTTGGGCCTG	2040
Db		1981		
QY	2041	2041	TTTAGGTACCTGGGTCTTTACCATAACCTTCGCATATATGATACAGTACCTTCTAAAGTGG	2100
Db		2041		
QY	2101	2101	CTTTGAATTGGAACTCTACAGTATGCACGAGTACTATTACATATATTGGTATCTCTCTGA	2160
Db		2101		
QY	2161	2161	ATTCTTTACGCATGGTTGATGTCAACATTTGAGTCGTGCCGATGGCTCTCAAATGGCAGA	2220
Db		2161		
QY	2221	2221	GGAAAGGATAAATGGAAGAGCAGCAGAAAGGCCGTAGTAAAGTAAACAAAGAAAAAAA	2280
Db		2221		
QY	2281	2281	GAAAGTTGCGCCATTGAGCGAGAGATCACAATGAGCAAGCATATCAGAACATGTGTGC	2340
Db		2281		
QY	2341	2341	TGGAATGTTTAAACCATTGGTAGCATTTTGACATGGACGGCAAAAGTACGTAAACCGAAGTT	2400
Db		2341		
QY	2401	2401	TGAGCTTGATAGTGAAACAAGTTCCGGTATGAAACACAGTTTGCTCCATTCAACAGTGTGAT	2460
Db		2401		
QY	2461	2461	GACCCCGCCCGCAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAG	2520
Db		2461		
QY	2521	2521	CCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACCTTTCACAGGCCAAA	2580
Db		2521		
QY	2581	2581	AATGATATTGGAAATATTCTTAACCCGGACCATGAGGTTAATAGAAATTTTAAAGGTTGC	2640
Db		2581		
QY	2641	2641	CAAAACCCAACTTTGTGGTTATGAAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGT	2700
Db		2641		
QY	2701	2701	TCCTCCTGAATTTGAATTTCTGTCTATAAATATTTCTGTGTGAAACTGTTTGAGA	2760
Db		2701		
QY	2761	2761	GAGACTGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTTCAGACCCCACTCTTAGAGGG	2820
Db		2761		
QY	2821	2821	CACATCACCAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAAACAACCTCATTA	2880
Db		2821		
QY	2881	2881	TAAGGAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGTTTTAAAG	2940
Db		2881		
QY	2941	2941	TGGTTTATTATGTTCCATGGAAGAACTCGTCTTATTTGAATGCATTTGATGAACGTTTATAT	3000
Db		2941		
QY	3001	3001	GGTTTTATTACAGATTTAATCACAAATCAATTTTTTATGAATGATTGAGTGAAAAATAGTGT	3060
Db		3001		
QY	3061	3061	TTATAAAGGTTAATAAATTTCTTGACAAAAA	3096

Db 3061 TTATAAAGGTTAATAATTTCTTGACAAAAAAA 3096  
|||||

RESULT 2  
AX537452 LOCUS AX537452 2488 bp DNA linear PAT 23-NOV-2002  
DEFINITION Sequence 23 from Patent WO02070709.  
ACCESSION AX537452  
VERSION AX537452.1 GI:25269238  
KEYWORDS .  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS Lal, P.G., Baughn, M.R., Yao, M.G., Wallia, N.K., Elliot, V.S., Xu, Y., Honchell, C.D., Yue, H., Ding, L., Gietzen, K.J., Ison, C.H., Lu, D.A., Hafalia, A.J., Ghandi, A.R., Thangavelu, K., Sanjanwala, M.M., Tang, Y.T., Ramkumar, J., Griffin, J.A., Swarnaker, A., Azimzai, Y., Sapperstein, S.K., Burford, N., Lee, E.A., Lu, Y., Tran, U.K. and Marquis, J.P.  
TITLE Molecules for disease detection and treatment  
JOURNAL Patent: WO 02070709-A 23 12-SEP-2002;  
INCYTE Genomics, Inc. (US)

FEATURES  
source Location/Qualifiers  
1..2488  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 7111920CB1"

BASE COUNT 768 a 461 c 567 g 692 t

ORIGIN

Query Match 78.1%; Score 2417.2; DB 6; Length 2488;  
Best Local Similarity 98.5%; Pred. No. 0;  
Matches 2470; Conservative 0; Mismatches 13; Indels 25; Gaps 2;

QY 474 GGCGGGCCGAGCGCGTCGTATTTCGGTCCGGACAGTGGCGCGCGGGT 533  
|||  
Db 1 GGCGGGCCGAGCGCGTCGTATTTCGGTCCGGACAGTGGCGCGCGGGT 60  
|||

QY 534 GACCACGGAGAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 593  
|||  
Db 61 GACCACGGAGAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 120  
|||

QY 594 ATGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAACAATCAAACTGGGTGGACATTAC 653  
|||  
Db 121 ATGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAACAATCAAACTGGGTGGACATTAC 180  
|||

QY 654 CCAAGATTTGAAGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCT 713  
|||  
Db 181 CCAAGATTTGAAGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCT 240  
|||

QY 714 ATTGGTCTTTTGAAGCCATGTCTGCTATTGAAATGATGGATCCCAGATGGATGCTGG 773  
|||  
Db 241 ATTGGTCTTTTGAAGCCATGTCTGCTATTGAAATGATGGATCCCAGATGGATGCTGG 300  
|||

QY 774 CATGATTGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 833  
|||  
Db 301 CATGATTGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 360  
|||

QY 834 CACTATTAAAAATTAAGATCTCACCTTGCCCTGAACCTGATAGGGATTATGGATACATGTTT 893  
|||  
Db 361 CACTATTAAAAATTAAGATCTCACCTTGCCCTGAACCTGATAGGGATTATGGATACATGTTT 420  
|||

QY 894 TTGCTGTTTTGATAACGTTGGTTAGAGGCCATTCACTGGCACAGACAGTATTTACGTGCTT 953  
|||  
Db 421 TTGCTGTTTTGATAACGTTGGTTAGAGGCCATTCACTGGCACAGACAGTATTTACGTGCTT 480  
|||

QY 954 TTACATTCATAATCCAGACTTTTATAGAAGATCCTGCTGATGAAGGCTTTTGTCTCTGGGAAT 1013  
|||  
Db 481 TTACATTCATAATCCAGACTTTTATAGAAGATCCTGCTGATGAAGGCTTTTGTCTCTGGGAAT 540  
|||

QY	1014	CTTGAAAATCTGTGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTTGAAGAGGA	1073
Db	541	CTTGAAAATCTGTGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTTGAAGAGGA	600
QY	1074	AGATTTTCAGTCAATGACTTATGGATTTTAAAAATGGTTAAACAGTGTGACAGATCTTCGAGT	1133
Db	601	AGATTTTCAGTCAATGACTTATGGATTTTAAAAATGGCTTAAACAGTGTGACAGATCTTCGAGT	660
QY	1134	TACAGGCATGTAAAGATGTGGAGGATGCACATGCAAGAAGAGTAAAGACTACTCGAAG	1193
Db	661	TACAGGCATGTAAAGATGTGGAGGATGCACATGCAAGAAGAGTAAAGACTACTCGAAG	720
QY	1194	TCGACAAAGGAGAAAGAGATCCAGAAAGTTGAACTAGAACACCAACAATGTTTAGCAGT	1253
Db	721	TCGACAAAGGAGAAAGAGATCCAGAAAGTTGAACTAGAACACCAACAATGTTTAGCAGT	780
QY	1254	ATTCAGCAGAGTGAATTTACTCGTGTGTTACTGACAGTGCCTTATAGCCCTTTACTAAGAA	1313
Db	781	ATTCAGCAGAGTGAATTTACTCGTGTGTTACTGACAGTGCCTTATAGCCCTTTACTAAGAA	840
QY	1314	AGAGACCAGTGCTGTTGCAGAAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTCTTTC	1373
Db	841	AGAGACCAGTGCTGTTGCAGAAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTCTTTC	900
QY	1374	TGCCATTCAATAATTCATTGCGATCATGGCATCCAGGCCCAGAAATGATACTACAAAAGGAGA	1433
Db	901	TGCCATTCAATAATTCATTGCGATCATGGCATCCAGGCCCAGAAATGATACTACAAAAGGAGA	960
QY	1434	TCATCCAATTATGATGGGTTTTGAACCCCTTGTGAAACCAGAGGCTACTTCCACCTACCTT	1493
Db	961	TCATCCAATTATGATGGGTTTTGAACCCCTTGTGAAACCAGAGGCTACTTCCACCTACCTT	1020
QY	1494	CCCTCGATATGCAAAAATAATTAAAAGGGAAGAAATGGTGAACTATTTTGCAGATTAAT	1553
Db	1021	CCCTCGATATGCAAAAATAATTAAAAGGGAAGAAATGGTGAACTATTTTGCAGATTAAT	1080
QY	1554	AGATAGAATAAAAACTGTCGTGAGGTTGTGAATTTAACAAATTTACATTTGATCCTGGA	1613
Db	1081	AGATAGAATAAAAACTGTCGTGAGGTTGTGAATTTAACAAATTTACATTTGATCCTGGA	1140
QY	1614	TTTTTCTGTGAATTTAGTGAACAGTCAACCATGTGTTCTTTCAAGATCTCTGTTACAAAAC	1673
Db	1141	TTTTTCTGTGAATTTAGTGAACAGTCAACCATGTGTTCTTTCAAGATCTCTGTTACAAAAC	1200
QY	1674	CACCTTCTCGTGGATACAAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAA	1733
Db	1201	CACCTTCTCGTGGATACAAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAA	1260
QY	1734	AGATGCACCTTCGGTCTTTTGCAGATCCTCCGAGTGCCTTCCCCCAAGTGCCTACCTATAT	1793
Db	1261	AGATGCACCTTCGGTCTTTTGTGAGTCCCTCCG--GTGCTTCCCCCAAGTGCCTACCTATAT	1318
QY	1794	AATAATCACCAGGCTAAGGACTGTATCGACTCCCTTTGTACTCACGTGTGTTCGGCCATTC	1853
Db	1319	AATAATCACCAGGCTAAGGACTGTATCGACTCCCTTTGTACTCACGTGTGTTCGGCCATTC	1378
QY	1854	TGTAGTCTTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAT	1913
Db	1379	TGTAGTCTTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAT	1438
QY	1914	ATTCTTGAGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTTAATAGGCAG	1973
Db	1439	ATTCTTGAGGAATTTGCCACCTTGCAGGATG-----AGGCAG	1475
QY	1974	AGAAGGTTGATGCAGCGCTTCACACCATGTCTGTTGAAACAGGAACCCCAAGGCAACATT	2033
Db	1476	AGAAGGTTGATGCAGCGCTTCACACCATGTCTGTTGAAACAGGAACCCCAAGGCAACATT	1535
QY	2034	TGGCTGTTTAGGTACCTGGGTCCTTTACCATAACCTTCGCATTATGATACAGTACCTTC	2093
Db	1536	TGGCTGTTTAGGTACCTGGGTCCTTTACCATAACCTTCGCATTATGATACAGTACCTTC	1595

Qy	2094	TAAGTGGCTTTGAATTTGGAACCTCTACAGTATGCACGAGTACTATTACATATATTGGTATC	2153
Db	1596	TAAGTGGCTTTGAATTTGGAACCTCTACAGTATGCACGAGTACTATTACATATATTGGTATC	1655
Qy	2154	TCTCTGAATTCCTTTACGCAATGGTTGATGTCAACATTTAGTCTGTCCTGATGCTCTCAAA	2213
Db	1656	TCTCTGAATTCCTTTACGCAATGGTTGATGTCAACATTTAGTCTGTCCTGATGCTCTCAAA	1715
Qy	2214	TGGCAGAGGAAAGGATAATGGAAGAGCAGCAGAAAGCCGTAGTAGTAAAAAAAACAAGA	2273
Db	1716	TGGCAGAGGAAAGGATAATGGAAGAGCAGCAGAAAGCCGTAGTAGTAAAAAAAACAAGA	1775
Qy	2274	AAAAAAGAAAGTTTCGCCATTGAGCCGAGAGATCACAATGAGCCAAGCATATCAGAACA	2333
Db	1776	AAAAAAGAAAGTTTCGCCATTGAGCCGAGAGATCACAATGAGCCAAGCATATCAGAACA	1835
Qy	2334	TGTGTGCTGGAATGTTTAAAAACCATGGTAGCATTTGACATGGACGGCAAAAGTACGTAAAC	2393
Db	1836	TGTGTGCTGGAATGTTTAAAAACCATGGTAGCATTTGACATGGACGGCAAAAGTACGTAAAC	1895
Qy	2394	CGAAGTTTGAGCTTGATAGTGAACAAGTTCGGTATGAACAACAGGTTTGTCTCCATTCAACA	2453
Db	1896	CGAAGTTTGAGCTTGATAGTGAACAAGTTCGGTATGAACAACAGGTTTGTCTCCATTCAACA	1955
Qy	2454	GTGTGATGACCCCGCCAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATA	2513
Db	1956	GTGTGATGACCCCGCCAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATA	2015
Qy	2514	AATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACTTTCAAC	2573
Db	2016	AATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACTTTCAAC	2075
Qy	2574	AGGCAAAAATGATATTGGAAAAATATTCCCTAACCCGGACCATTGAGGTTAATAGAAATTTAA	2633
Db	2076	AGGCAAAAATGATATTGGAAAAATATTCCCTAACCCGGACCATTGAGGTTAATAGAAATTTAA	2135
Qy	2634	AGGTTGCCAAACCCCACTTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAGGAAT	2693
D'	2136	AGGTTGCCAAACCCCACTTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAGGAAT	2195
Qy	2694	CTAAAGTTCCTCCTGAAATTTTGATTTCTCTGCTCATAAATATTTTCTGTTGTGAAACTTG	2753
Db	2196	CTAAAGTTCCTCCTGAAATTTTGATTTCTCTGCTCATAAATATTTTCTGTTGTGAAACTTG	2255
Qy	2754	TTTGAGAGAGACTGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTTCAGACCCAACTCT	2813
Db	2256	TTTGAGAGAGACTGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTTCAGACCCAACTCT	2315
Qy	2814	TAGAGGGCACATCACAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAACAA	2873
Db	2316	TAGAGGGCACATCACAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAACAA	2375
Qy	2874	CTCATTATAAGGAATACCTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGT	2933
Db	2376	CTCATTATAAGGAATACCTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGT	2435
Qy	2934	TTTAAAGTGGTTTTATTATGTTCCATGGAAGAAACTGGTCTTTATTGAAT	2981
Db	2436	TTTAAAGTGGTTTTATTATGTTCCATGTAAGACACTGGGTTCCATTAAAT	2483
RESULT 3			
AY102701			
LOCUS			
DEFINITION	Mus musculus embryonic growth-associated protein EGAP mRNA, complete cds.		
ACCESSION	AY102701		
VERSION	AY102701.1		
KEYWORDS	GI:21539895		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		



Db 1575 CTTTGAGCTGGAGCTGTACAGCATGCATGAGTACTACTACATCTACTGGTACCTCTCCGA 1634

Qy 2161 ATTCCCTTTACGCATGGTTGATGTCAACATTGAGTCGTGCCGATGGCTCTCAAAATGGCAGA 2220

Db 1635 GTTCCTGTATGCATGGCTGATGTCAACCCTGAGCCGCGTACGGCTCTCAGATGGCAGA 1694

Qy 2221 GGAAAGGATAATGGAAGAGCAGCAGAAAAGCCGTAGTAGTAAAAAACAAGAAAAAAA 2280

Db 1695 GGAGAGGATCATGGAGGAACAGCAGAAAGCCGCAGCAGCAAAAAAACAAGAGAAAAA 1754

Qy 2281 GAAAGTTCGCCCATTTGAGCCGAGAGATCAAAATGAGCCAGCATATCAGAACATGTGTGC 2340

Db 1755 GAAAGTTCGTCCTGTTGAGCCGAGAGATCAAAATGAGCCAGGCTATCAGAACATGTGTGC 1814

Qy 2341 TGGAATGTTTAAAAACCATGGTAGCAATTTGACATGGACGGCAAGTACGTAAACCGAAGTT 2400

Db 1815 TGGGATGTTCAAGACGATGGTTCGCTTTTGACATGGATGGCAAGTGCGAAGCCCAAGTT 1874

Qy 2401 TGAGCTTGATAGTGAACAAAGTTCGGTATGAACACAGGTTTGCTCCATTCAACAGTGTGAT 2460

Db 1875 TGAGCTTGACAGTGAGCAGGTCCGCTATGAGCACAGGTTTGACCGTTCAACAGTGTGAT 1934

Qy 2461 GACCCGCGCCAGTGCACTACTTACAGTTCAAGGAAATGCTGCACCTCAATAAATATAG 2520

Db 1935 GACGCGCCACCACTGCACTACCTGAGTTCAAGGAAATGCTGCACCTCAATAAATACAG 1994

Qy 2521 CCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCATA 2580

Db 1995 CCCTCCTCCTCAGTCTCCAGAGCTGTATGTGGCAGCTAGTAAGCACTTTTCAGCAGACAA 2054

Qy 2581 AATGATATTGGAAAAATATTCCCTAACCCGGACCATGAGGTTAATAGAAATTTAAAGTTGC 2640

Db 2055 AATGATCCTTGAAAAATATCCCAATCCAGACCGTGAGGTCAGTAGAAATCCTGAAGTTGC 2114

Qy 2641 CAAACCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGAGTCAAAGGT 2700

Db 2115 CAAGCCAAACTTTGTGGTTATGAAGCTTTAGCAGGAGGACACAAAAAGAGTCAAAGGT 2174

Qy 2701 TCCTCCTGAATTTGATTTCTCTGCTCAFAAATATTTTCTGTTGAAAACTTGTGTTGAGA 2760

Db 2175 TCCTCCCGAATTCGACTTCTCTGTTCAAAAATATTTTCTGTTGAAAACTGTTTGAAA 2234

Qy 2761 GAGACTGGGAGGTGGCCATAAAGGGCAGAGTCTTCTTTTCAGACCCAACTCTTAGAGGG 2820

Db 2235 GAGACTGCAAAAGATGACCGTGTCCACTC-----TAGGGAA 2269

Qy 2821 CACATCACCAAGGCTCCACATCACG--GGAAGTGAGATGGATTTCTTGGGTAACAACATCAT 2878

Db 2270 GACATCAGTCTGCCACACATGGTGACAGAAAGTGATCTCTTGAATGACAGACAGCTTGC 2329

Qy 2879 TATAAGGAATACTTTTAGTTTGACAGCCCTTATATGACATGAATGAAAACTGCTGTTTAA 2938

Db 2330 TACAGGAAGACTTTTCAGTTTACCACCCCTTACCTAGAGAGGATGAGAACTGCTGTTTAA 2389

Qy 2939 AGTGGTTTATATGTTCCATGGAAGAAACTGGTCTTATTGAATGCATTGATGAACGTTAT 2998

Db 2390 AGTGGTTTATGTACTCCATGGGTGTAACGTG---CTGGAATGCATTGGTGAACGCTAC 2445

Qy 2999 ATGTTTTTATTACAGATTTAATCAAAATCAATTTTATGAATGATGTAGTGAAAAATAGT 3058

Db 2446 ATGGTTTGAATTACAGACTTAATTTGTAACAAATTTTAAATGAATGATGTGTG-AAATAGT 2504

Qy 3059 GTTTATAAAGGTTAATAAATTTCTTGACAAAAA 3096

Db 2505 GTTTGTAAAGGTTAATAAATTTCTTGACAAAAA 2542

RESULT 4  
AF272892 AF272892 2543 bp mRNA linear ROD 15-SEP-2000  
LOCUS Rattus norvegicus corneal wound healing related protein mRNA,  
DEFINITION complete cds.

ACCESSION AF272892  
VERSION AF272892.1 GI:8926319  
KEYWORDS  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 2543)  
AUTHORS Yi,X.J., Li,X.F. and Yu,F.S.  
TITLE A novel epithelial wound-related gene is abundantly expressed in  
developing rat cornea and skin  
JOURNAL Curr. Eye Res. 20 (5), 430-440 (2000)  
MEDLINE 20314365  
PUBMED 10855038  
REFERENCE 2 (bases 1 to 2543)  
AUTHORS Li,X. and Yu,F.  
TITLE Direct Submission  
JOURNAL Submitted (26-MAY-2000) The Schepens Eye Research, 20 Staniford  
Street, Boston, MA 02114, USA  
FEATURES  
Location/Qualifiers  
1..2543  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
64..2241  
/codon\_start=1  
/product="corneal wound healing related protein"  
/protein\_id="AAF81791.1"  
/db\_xref="GI:8926320"  
/translation="MVMAAIVDDDDASGWELNVPKMEKSSTSWVDITQDFEDACRELK  
LGELLHDKLFGLEAFMSAIEEMDPKMDAGMIGNVNRKVLNFEQAVKDGTIKIKDLSL  
PELIGIMDTCFCLITWLEHSLAQVTFTCLYIHNPDFIEDPAMKAFALGILKICDIA  
REKVNKAAAFEEEDFQSMYGFKMAANGVTDLRVTGMLKDVEDDMQRRVKSTRSQGEE  
RDPEVELEHQOCLAAFSRVKFTRVLLTVLIAFTKETSAAVEAQKLMVQAADLLSAIH  
TSLHHGIQAQNGTTXGDHPIMMGFEPLVNQRLLPPTFPFYAKIIKREEMVNYFSRLID  
RIKTVCEVNVNLPNLHCILDFFCEFESEQSPCVLSRLLOTTFVLVDNKKVFGTHLMQDMV  
KDAPRSFVSPVLPKCLYNNHQAQKDCIDSFVTHCVRPFCSLVQIIGHNRARQDKL  
GHILEEFATLQDEAEKVDAALHTMLLKQEPQORQHLACLGLTWVLYHSLRIMIQLLSGF  
DLELYSMHEYVYIYWLSEFLYAWLMSTLSRADGSOAABERIMEEQOKGRSSKTKKK  
KKVRPLSREITMSQAYQNCAGMKMTMAFMDGKVRPKFELDSQVRYEHRFAPFN  
SVMTPPVHYLQFKEMSDLSKYSPRPQPPPELYVAASGHFQQAAMILESIPNADREVS  
ILKVAKNFNVVMKLLAGHHKSKVPPPEFDFSVHKYFPVVKLV"

BASE COUNT 725 a 554 c 627 g 637 t  
ORIGIN

Query Match 58.1%; Score 1799.2; DB 10; Length 2543;  
Best Local Similarity 84.0%; Pred. No. 0;  
Matches 2139; Conservative 0; Mismatches 343; Indels 64; Gaps 7;

Qy 550 GGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTACAGGATGGGAGCTCAGTATG 609  
|||||  
Db 58 GGCATAATGGTTATGAAAGCCGACGTAGATGATGACGCTTCGGGATGGGAGCTCAATGTC 117  
|||||

Qy 610 CCAGAAAAAATGGAGAAAAGCAATACAAAACCTGGGTGGACATTACCCAAAGATTTTGAAGAA 669  
|||  
Db 118 CCTGAGAAAAATGGAAAAAGTAGCACAAAGCTGGGTGGACATAACCAGGACTTTGAAGAT 177  
|||||

Qy 670 GCTTGTGAGAAATTAAAGTTGGGAGAACTACTTTCATGATAAGCTATTTGGTCTTTTGA 729  
|||||  
Db 178 GCTTGTGAGAGCTGAAGTTGGGAGAACTGCTTCAGATAAGCTGTTGGTCTGTTTGA 237  
|||||

Qy 730 GCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAACCAA 789  
|||||  
Db 238 GCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCCGGATGATCGGAAACCAA 297  
|||||

Qy 790 GTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAAAATAAA 849  
|||  
Db 298 GTGAATAGAAAAAGTCCTCAATTTTGAACAAGCCGTCAGGATGGCACCATTAAAAATAAA 357  
|||||

Qy 850 GATCTCACCTTGCCCTGAACGTATAGGATTTATGGATACATGTTTCTGCTGTTGATAACG 909  
|||  
Db 358 GACCTCAGCTTGCCCTGAACGTATAGGAATAATGGATACCTGTTTCTGCTGTTGATCACA 417  
|||||





Db 1141 AAAAACTGCTGTGAGGTTGTGAATTTAACAAATTTACATTGTATCCTCGGATTTTTTCTG 1200

QY 1623 TGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTCCT 1682

Db 1201 TGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTCCT 1260

QY 1683 GGTGGATAACAAAGGTCTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACT 1742

Db 1261 GGTGGATAACAAAGGTCTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACT 1320

QY 1743 TCGGTCTTTTGTGAGATCCTCCGAGTGTCTTCCCCAAAGTGCTACCTATATAATAATCAC 1802

Db 1321 TCGGTCTTTTGTGAG--TCCTCCGGTGTCTTCCCCAAAGTGCTACCTATATAATAATCAC 1378

QY 1803 CAGGCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTTCGGCCATTTCTGTAGTCTT 1862

Db 1379 CAGGCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTTCGGCCATTTCTGTAGTCTT 1438

QY 1863 ATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTTCATATTCTTGAG 1922

Db 1439 ATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTTCATATTCTTGAG 1498

QY 1923 GAATTTGCCACCTTCAGGATGAGTTTATGACATTTTATTTAATAGGCAGAGAAGGTTG 1982

Db 1499 GAATTTGCCACCTTCAGGATG-----AGGCAGAGAAGGTTG 1535

QY 1983 ATGCAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATTTGGCCTGTT 2042

Db 1536 ATGCAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATTTGGCCTGGT 1595

QY 2043 TAGGTACCTGGTGCTTTTACCATAACCTTCGCAATATGATACAGTACCTTCTAAGTGGCT 2102

Db 1596 TAGGTACCTGGTGCTTTTACCATAACCTTCGCAATATGATACAGTACCTTCTAAGTGGCT 1655

QY 2103 TTGAATTGGAACCTACAGTATGCACGAG-TACTATTACATATATTGGTATCTCTCTGAA 2161

Db 1656 TTGAATTGGAACCTACAGTATGCACGAGATACTATTACATATATTGGTATCTCTCTGAA 1715

QY 2162 TTCCTTTACGCATGTTGATGTCAACATTGAGTCGTCCGATGGCTCTCAATGGCAGAG 2221

Db 1716 TTCCTTTACGCATGTTGATGTCAACATTGAGTCGTCCGATGGCTCTCAATGGCAGAG 1775

QY 2222 GAAAGGATAATGGAAGACGACGAGAAAGGCCGTAGTAGTAAAAAACAACAAAAAAG 2281

Db 1776 GAAAGGATAATGGAAGACGACGAGAAAGGCCGTAGTAGTAAAAAACAACAAAAAAG 1835

QY 2282 AAA 2284

Db 1836 AAA 1838

RESULT 6

BC053286

LOCUS

DEFINITION Danio rerio cDNA clone MGC:64157 IMAGE:6797204, complete cds.

ACCESSION BC053286

VERSION BC053286.1 GI:31418975

KEYWORDS MGC.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 2467)

REFERENCE

AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,

McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettaman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 2467)

Strausberg,R.

Direct Submission

Submitted (02-JUN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Leonard I. Zon, M.D.

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 117 Row: n Column: 11

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenomeScan gene prediction, Similarity but not identity to protein.

FEATURES

source

1. .2467

/organism="Danio rerio"

/mol\_type="mRNA"

/db\_xref="taxon:7955"

/clone="MGC:64157 IMAGE:6797204"

/tissue\_type="Kidney, zebrafish"

/clone\_lib="NCI CGAP\_ZKId1"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6.1"

58. .2232

/codon\_start=1

/product="Unknown (protein for MGC:64157)"

/protein\_id="AAH53286.1"

/db\_xref="GI:31418976"

/translation="MVVKSSVEEEGGWGLGIPKMRNNANWVDVTQEFKGACKELKL GELLHDKLFGLEAMSALIEEMDPKMDAGMIGNVNRKVLNFDQAVKDEAIRVKDLSIP ELIGIMDTCCCLITWLEHSLAQTVFTCLYVHPDLIQDPALKAPALGILKICDIAR EKVNKAAVFEEEDFQAMTYGFKMANNVTDLRVTGMLKDVEDELQPKVKSRSRQGEOR DPEVELDHQQCLALFSRVKFTRLLSALISFTKKTSAVSEAQKLSQAADLLPAVHA TIQYGIQSONDTTKGDHPIMMGFEPLVNQRLLPPTFPFYAKIKREEMVNYFSKLIER IKSVCEVINITNLHSILDFFCFESEQSPCVLSRSLQTTFLIDNKKVFGTHLMQDMIK DALRCFVSPPLVSSKCSLNNHQAQKDYDSFVTHCTRPFCSLIQIHGHNARQRDKLG HILEEFATLQDEAEKVDAALHGLMKLEPQRQHLACLGTWILYHNLIRIMIQYLLSGFE LELYSMHEYIYIYWLSEFLYAWLMSLTSRADSSQMAEERILEEQLKVRSSKSKKKK KARPLSKEITMSQAYQNMCMAGYKMTWIALDMDRKVRKPKQFELDSEQVRYEHRFAPFNS VVTPPVHYTIQFKEMSDLKYNPPPSADLYMAASKHFQQAKLLENVTSPDAEVNRI LKVAKPNIIVMKLLAGCHKKETKALPEPDFSAHKYFPPIVKIL"

CDS

BASE COUNT 720 a 574 c 600 g 573 t

ORIGIN

Query Match 40.9%; Score 1264.8; DB 5; Length 2467;

Best Local Similarity 74.3%; Pred. No. 8.3e-276; Matches 1648; Conservative 0; Mismatches 542; Indels 28; Gaps 3;									
QY	543	AGAA	GTAGGC	ATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGGATGGAGCT	602				
Db	45	AAA	AGCTGTCTCCATGGTGATGAAGTCATCGGTGAGGAGGAGGAGGATGGGTCT	104					
QY	603	CAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAAACTGGGTGGACATTACCCCAAGATTT	662						
Db	105	TGGGATCCCAGAGAAAGAT---GAGAAACAATGCCAACTGGGTGGACGTCACCCAGGAGTT	161						
QY	663	TGAAGAAGCTTGTGAGAAATTAAAGTTGGGAGAACTACTTTCATGATAAGCTATTTGGTCT	722						
Db	162	TAAAGGAGCTTGTAAGAACACTCAAACTTGGCGAGCTGCTCCATGACAAGCTATTCGGTCT	221						
QY	723	TTTTGAAGCCCATGTCGTATGTAAATGATGGATCCCAAGATGGATGCTGGCATGATTGG	782						
Db	222	GTTGGAAGCCCATGTCGCCCATGTAGATGATGGATCCGAAGATGGATGCTGGAATGATCGG	281						
QY	783	AAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAA	842						
Db	282	AAATCAAGTCAACCGTAAAGTCTCAACTTTGACCAAGCTGTTAAGGATGAAGCCATACG	341						
QY	843	AATPAAAGATCTCACCTTGCCCTGAACTGATAGGGATTATGGATACATGTTTTTGCCTGTTT	902						
Db	342	GGTGAAGATCTAAGTATTCCTGAGTTCATCGGGATCATCGGATCATGACACATGTTTCTGCTGCT	401						
QY	903	GATAACGTGGTTAGAAAGGCCATTCACCTGGCACAGACAGATATTACGTGCCTTTACATTCA	962						
Db	402	GATCACGTGGCTGGAGGGCAITCTCTGGCTCAGACCGTCTTTACATGTCTGTACGTTCA	461						
QY	963	TAATCCAGACTTTATAGAAGATCCCTGCTATGAAGGCTTTTGTCTTGGGAATCTTTGAAAAT	1022						
Db	462	TAATCCGGATCTGATTCAAGACCCCTGCGCTCAAAGCCITTCGTCTGGGCATCCTGAAAAT	521						
QY	1023	CTGTGACATTTGCAAGGGAAAAAGTAAATAAAGCTGCTGTTTGAAGAGGAAGATTTTCA	1082						
Db	522	CTGTGACATCGCCCGTGAAAAAGTCAACAAAGCTGCCGTGTTTGAAGAGGAAGATTTTCA	581						
QY	1083	GTCAATGACTTATGGATTTTAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGGCAT	1142						
Db	582	GGCATGACTTACGGCTTTTAAATGGCCAAACAACGTAACCTGATCTGAGAGTGACGGGAAT	641						
QY	1143	GCTAAAAAGATGTGGAGGATGACATGCAAAAGAAAGAGTAAAGAGTACTCGAAGTCGACAAAG	1202						
Db	642	GCTGAAAGATGTTGAGGATGAACCTTCAGAGGAAAAAGTAAAGAGTACCCGACGTCGACAGGG	701						
QY	1203	AGAGAAAAAGATCCAGAAAGTTGAAGTACTAGAACCAACAACAAATGTTTAGCAGTATTCAGCAG	1262						
Db	702	TGAACACGGAGATCCTGAGGTGGAAGTGGATCATCAGCAGTGCTTGGCTCTGTTTCAGTCG	761						
QY	1263	AGTGAATTTACTCGTGTGTTTACTGACAGTGTCTTATAGCCCTTTACTAAGAAAGAGACCAG	1322						
Db	762	GGTTAAATTCACCTCGTCTGCTGCTCAGCGCGCTAATCTCCTTCACCAAAAAAGAGACGAG	821						
QY	1323	TGCTGTGCAGAAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCA	1382						
Db	822	TGCTGTGAGCGAAGCACAGAAGCTGATGAGTCAAGCTGCTGATTTACTGCTGCAGTCCA	881						
QY	1383	TAATTCAATTGCATCATGGCATCCAGGCCAGAATGATACTACAAAAGGAGATCATCCAAAT	1442						
Db	882	TGCTACCATAACAGTACGGCATACAGTCGCAGAACGACACTACAAAGGAGATCACCCGAT	941						
QY	1443	TATGATGGGTTTTGAACCCCTTGTGAACCAAGAGGCTACTTCCACCTACCTTCCCTCGATA	1502						
Db	942	CATGATGGGTTTTGAGCCCTCTGGTTAAACCAAGCGCCTGCTCCACCTACGTTCCCCCGATA	1001						
QY	1503	TGCAAAAAATAATTAAGGGAAGAAATGGTGAACATAATTTTGCAAAGATTAATAGATAGAAAT	1562						
Db	1002	TGCCAAGATCATCAAGCGAGGAGATGGTCAACTACTTCAGCAAAACTCATTTGAGCGCAT	1061						
QY	1563	AAAACTGCTGTGAGGTTGTGAATTTAAACAAATTTTACATTTGATCCTGGATTTTTTCTG	1622						

Dr	1062	TAAG	AGCGTGTGTGAGGT	CATCAACA	CATCACCA	ACCTGCACAGCATCCTGGATTTCTTCTG	1121
QY	1623	TGAATTTAGTGAACAGATCACCATGTGTTCTTTCAAGATCTCTGTTACAAAACCACTTTTCC	1682				
Db	1122	TGAATTCAGTGAGCAGTCTCCATGTGTGCTGTCTAGATCTCTACTGCAGACCAACGTTTCT	1181				
QY	1683	GGTGGATAACAAAAAGGTCTTTTGGAACTCATCTCATGCAAGACATGGTGAAAAGATGC	1742				
Db	1182	GATCGACAACAAGAGGTGTTTGGGACTCATCTGATGAGGATATGATTAAGATGCGCT	1241				
QY	1743	TCGGTCTTTTGTGAGATCCTCCGAGTGTCTTCCCCCAAGTGTACCTATATAATAATCAC	1802				
Db	1242	GCGCTGTTTTGTGAG--TCCACCTGTTCTTTCCCTCAAAATGCAGTTTGAATAATAACCA	1299				
QY	1803	CAGGCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTCGGCCATTTCTGTAGTCTT	1862				
Db	1300	CAGGCTAAAGATTATATCGACTCCTTTGTTCACACACTGCACGCGGCCGTTCTG	1359				
QY	1863	ATTGAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAG	1922				
Db	1360	ATCCAGATTACGGACACAAACCGAGCCGACAGAGAGACAAACTGGGTACATTTCTGGAG	1419				
QY	1923	GAATTTGCCACCTTGCAGGATGAGTTTATGACATTTTATAGGCAGAGAAGGTTG	1982				
Db	1420	GAGTTCGCCACACTTCAGGATG-----AGGCAGAGAAGGTGG	1456				
QY	1983	ATGCAGCGCTTCACACCAATGCTGTTGAAACAGGAACCCCAAGGCAACATTTGGCCCTGTT	2042				
Db	1457	ACGCTGCTCTTACCGTCTGTGATGAAGCTGGAGCCGCGGCGCAGCATTTAGCCTGTC	1516				
QY	2043	TAGGTACCTGGGTCTTTTACCATAAACCTTCGCATTTATGATACAGTACCTTCTAAGTGGCT	2102				
Db	1517	TGGGCACCTGGATCCTTACACAACCTGCGCATCATGATCCAGTACCTGCTCAGCGCT	1576				
QY	2103	TTGAATTGGAACCTCTACAGTATGCACGAGTACTATTACATATATTGGTATCTCTCTGAAT	2162				
Db	1577	TTGAGCTGGAGCTCTACAGCATGCACGAATACTATTACATCTACTGTTGTTGTTGTCAGAGT	1636				
QY	2163	TCCTTTACGCATGGTTGATGTCAACCATTGAGTGTGCGGATGGCTCTCAAATGGCAGAGG	2222				
Db	1637	TCTTGTATGCGTGGCTGATGTGACGTTGAGTCGAGCCGACAGCTCTCAGATGGCGGAGG	1696				
QY	2223	AAACGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAAAAACAAGAAAAAAGA	2282				
Db	1697	AGAGAAATCCTGGAGGAGCAGCTCAAAGTGCGCAGCAGCAAGAGAGCAAGAAAGA	1756				
QY	2283	AAGTTCGCCCATTTAGCCGAGAGATCACAAATGAGCCAAAGCATATCAGAAACATGTGTGCTG	2342				
Db	1757	AAGTCTGCTCCTCTGAGTAAAGAGATCACCATGAGCCAGGCTTACCAGAAACATGTGTGCTG	1816				
QY	2343	GAATGTTTTAAACCATGGTAGATTTGACATGGACGGCAAGTACGTAACCGAAGTTTG	2402				
Db	1817	GGATGTACAAAGACGATGATCGCGCTGGATATGGATCGAAAGGTGCGCAAAACCTCAGTTTG	1876				
QY	2403	AGCTTGATAGTGAACAAAGTTTCGGTATGAACAACAGGTTTGTCTCCATTCAACAGTGTGATGA	2462				
Db	1877	AGCTGGACAGTGAGCAGGTGCGCTACGAGCATCGCTTCGCCCTTTCAACAGTGTGGTCA	1936				
QY	2463	CCCGCCCGCAGTGCACCTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCC	2522				
Db	1937	CACCACCAACGATTCACATACAGTTTAAGGAAATGTCCGATTTGAAGAAGTACAATC	1996				
QY	2523	CTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAAGCACTTTCAACAGGCAAAA	2582				
Db	1997	CTCCTCCTCGCTCAGCGGATCTCTATATGCTGGCTGCCAGTAAACATTTCCAGCAGGCCAAAC	2056				
QY	2583	TGATATTGGAAAAATATTCCTAAACCCGGACCATGAGGTTAATAGAAATTTTAAAGTTGCCA	2642				
Db	2057	TCTTACTGGAAAAATGTCACAGCCCTGATGTCAGAGGTTAATTCGTATCCTTAAAGTGGCCA	2116				
QY	2643	AACCCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTC	2702				
Db	2117	AACCCAAATATTGTTGTAATGAAGCTGCTGGCTGGAGGACACAAGAAAGGAGACCAAGGCC	2176				

QY	2703	CTCC	GAATTTGATTTCTCTGCTCATAAATATTTTCCCTGTTGTGAAACTTGTTGAGA	2760
Db	2177	TTCCT	GAATTTGACTTCTCTGCTCACAAATACTTCCCATAGTCAAAATCCTATGAGA	2234
RESULT 7				
AK026296				
LOCUS				
DEFINITION Homo sapiens cDNA: FLJ22643 fis, clone HSI07031. linear PRI 29-SEP-2000				
ACCESSION AK026296				
VERSION AK026296.1 GI:10439112				
KEYWORDS oligo capping; fis (full insert sequence).				
SOURCE Homo sapiens (human)				
ORGANISM Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).				
FEATURES				
source				
1..997				
/organism="Homo sapiens"				
/mol_type="mRNA"				
/db_xref="taxon:9606"				
/clone="HSI07031"				
/tissue_type="human small intestine"				
/clone_lib="HSI"				
/note="cloning vector pME18SFL3"				
15..650				
/note="unnamed protein product"				
/codon_start=1				
/protein_id="BAB15435.1"				
/db_xref="GI:10439113"				
/translation="MHEYIYIYWILSEFLYAWLMSLTLSRADGSQMAERIMEEQQKGR SSKTKKKKKVRPLSREITMSQAYQNCAGMFKTMVAFDMDGKVRKPKFELDSQVRY EHRFAPNSVMTPPVHYLQFKEMSDLNKYSPPPQSPELXVAASKHFQQAKMILENIP NPDHEVNRILKVAKPNFVVMKLLAGGHKKESKVPPEFDFAHKYFPVVKLV"				
CDS				
BASE COUNT				
ORIGIN				
Query Match				
Best Local Similarity 31.9%; Score 988; DB 9; Length 997;				
Matches 988; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	2109	TGGA	ACTCTACAGTATGCACGAGTACTATTACATATATTGGTATCTCTCTGAATTCCTTT	2168
Db	1	TGGA	ACTCTACAGTATGCACGAGTACTATTACATATATTGGTATCTCTCTGAATTCCTTT	60
QY	2169	ACGC	ATGTTGATGTCAACATTGAGTCGTCGCCGATGGCTCTCAAATGGCAGAGAAAGGA	2228
Db	61	ACGC	ATGTTGATGTCAACATTGAGTCGTCGCCGATGGCTCTCAAATGGCAGAGAAAGGA	120
QY	2229	TAAT	GGAAGAGCAGCAGAAAGCCGCTAGTAGTAAAAAACAAGAAAAAAGAAAGTTC	2288

Db	121	TAATGGAAGAGCAGCAGAAAGCCGCTAGTAGTAAAAAACAAGAAAAAAGAAAGTTC	180
Qy	2289	GCCCATTTGAGCCGAGAGATCACAAATGAGCCCAAGCATATCAGAACATGTGTGCTGGAATGT	2348
Db	181	GCCCATTTGAGCCGAGAGATCACAAATGAGCCCAAGCATATCAGAACATGTGTGCTGGAATGT	240
Qy	2349	TTAAAACCATGGTAGCATTTGACATGGACCGCAAGTACGTAAACCGAAGTTTGAGCTTG	2408
Db	241	TTAAAACCATGGTAGCATTTGACATGGACCGCAAGTACGTAAACCGAAGTTTGAGCTTG	300
Qy	2409	ATAGTGAACAAAGTTTCGGTATGAACACACAGGTTTGCTCCATTCAACAGTGTGATGACCCCGC	2468
Db	301	ATAGTGAACAAAGTTTCGGTATGAACACACAGGTTTGCTCCATTCAACAGTGTGATGACCCCGC	360
Qy	2469	CGCCAGTGCCTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTC	2528
Db	361	CGCCAGTGCCTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTC	420
Qy	2529	CTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAGCACCTTTC AACAGGCCAAAAATGATAT	2588
Db	421	CTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAGCACCTTTC AACAGGCCAAAAATGATAT	480
Qy	2589	TGAAAAATATTCTTAACCCGGACCATTGAGGTTAATAGAAATTTTAAAGTTGCCAAACCCA	2648
Db	481	TGAAAAATATTCTTAACCCGGACCATTGAGGTTAATAGAAATTTTAAAGTTGCCAAACCCA	540
Qy	2649	ACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTTAAAGTTCCCTCCTG	2708
Db	541	ACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTTAAAGTTCCCTCCTG	600
Qy	2709	AATTTGATTTCTCTGCTCATAAATATTTTCTGTTGTGAAACTTGTGAGAGAGACTGG	2768
Db	601	AATTTGATTTCTCTGCTCATAAATATTTTCTGTTGTGAAACTTGTGAGAGAGACTGG	660
Qy	2769	GGAGTGGCCATAAAGGGGCAGAGTCTTCTTTCAGACCCCAACTCTTAGAGGGCACATCAC	2828
Db	661	GGAGTGGCCATAAAGGGGCAGAGTCTTCTTTCAGACCCCAACTCTTAGAGGGCACATCAC	720
Qy	2829	CAGGCTCCACATCACGGGAAGTGAGATGGATTCTTGGGTAAACAACCTCATTATAAGGAAT	2888
Db	721	CAGGCTCCACATCACGGGAAGTGAGATGGATTCTTGGGTAAACAACCTCATTATAAGGAAT	780
Qy	2889	ACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGTTTTAAAGTGCTTTAT	2948
Db	781	ACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGTTTTAAAGTGCTTTAT	840
Qy	2949	TATGTTCCATGGAAGAACTGGTCTTATTGAATGCATTGATGAACGTTTATATGGTTTAT	3008
Db	841	TATGTTCCATGGAAGAACTGGTCTTATTGAATGCATTGATGAACGTTTATATGGTTTAT	900
Qy	3009	TACAGATTTAATCAGAAATCATTTTTTATGAATGATTGAGTGAAAAATAGTGTATATAAG	3068
Db	901	TACAGATTTAATCAGAAATCATTTTTTATGAATGATTGAGTGAAAAATAGTGTATATAAG	960
Qy	3069	GTTAATAAATTTCTTGACAAAAA 3096	
Db	961	GTTAATAAATTTCTTGACAAAAA 988	
RESULT 8			
AK056059			
LOCUS	AK056059	1796 bp	mRNA
DEFINITION	Homo sapiens cDNA FLJ31497 fis, clone NT2NE2005317, highly similar to Rattus norvegicus corneal wound healing related protein mRNA.		
ACCESSION	AK056059		
VERSION	AK056059.1 GI:16551204		
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1		

AUTHORS Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T., Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K., Kamiyara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.  
NEDO human cDNA sequencing project  
Unpublished  
JOURNAL 2 (bases 1 to 1796)  
REFERENCE Isogai,T., Otsuki,T. and Sugiyama,T.  
AUTHORS Direct Submission  
TITLE Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
JOURNAL NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.  
FEATURES Location/Qualifiers  
source 1..1796  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2NE2005317"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2NE2"  
/note="cloning vector: pME18SFL3-mRNA from NT2 neuron after the differentiation of NT2 neuronal precursor cells."  
BASE COUNT 490 a 365 c 463 g 478 t  
ORIGIN  
Query Match 28.6%; Score 885.4; DB 9; Length 1796;  
Best Local Similarity 99.9%; Pred. No. 9e-190;  
Matches 886; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 549 AGGCATAATGGTTATGAAGCTTCTGTAGATGATGACGATTGAGGAGCTCAGTAT 608  
Db 345 AGGCATAATGGTTATGAAGCTTCTGTAGATGATGACGATTGAGGAGCTCAGTAT 404  
QY 609 GCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCCAAAGATTTTGAAGA 668  
Db 405 GCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCCAAAGATTTTGAAGA 464  
QY 669 AGCTTGTGAGAAATTAAGTTGGAGAACTACTTCATGATAAGCTATTTGGTCTTTTGA 728  
Db 465 AGCTTGTGAGAAATTAAGTTGGAGAACTACTTCATGATAAGCTATTTGGTCTTTTGA 524  
QY 729 AGCCATGCTCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAACCA 788  
Db 525 AGCCATGCTCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAACCA 584  
QY 789 AGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGATGGCACTATTAAAAATTAA 848  
Db 585 AGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGATGGCACTATTAAAAATTAA 644  
QY 849 AGATCTCACCTTGCCTGAACTGATAGGGATTATGGATACATGTTTGTGTTTGTATAAC 908  
Db 645 AGATCTCACCTTGCCTGAACTGATAGGGATTATGGATACATGTTTGTGTTTGTATAAC 704  
QY 909 GTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTTACGTGCTTTTACATTCATAATCC 968  
Db 705 GTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTTACGTGCTTTTACATTCATAATCC 764  
QY 969 AGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTCTGGGAATCTTGAAAAATCTGTGA 1028

Db 765 AGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAAAATCTGTGA 824  
QY 1029 CATTGCAAGGGAAAAAGTAGTAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTCAAT 1088  
Db 825 CATTGCAAGGGAAAAAGTAGTAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTCAAT 884  
QY 1089 GACTTATGGATTAAAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGCGATCTAAA 1148  
Db 885 GACTTATGGATTAAAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGCGATCTAAA 944  
QY 1149 AGATGTGGAGGATGACATGCACAAAGAGAGTAAAGAGTACTCGAAGTCGACAAAGGAGAAGA 1208  
Db 945 AGATGTGGAGGATGACATGCACAAAGAGAGTAAAGAGTACTCGAAGTCGACAAAGGAGAAGA 1004  
QY 1209 AAGAGATCCAGAAGTTGAACCTAGAACACCAACAATGTTTTCAGCAGTATTCAGCAGAGTGAA 1268  
Db 1005 AAGAGATCCAGAAGTTGAACCTAGAACACCAACAATGTTTTCAGCAGTATTCAGCAGAGTGAA 1064  
QY 1269 ATTTACTCGTGTGTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAGAGACCAGTGTGT 1328  
Db 1065 ATTTACTCGTGTGTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAGAGACCAGTGTGT 1124  
QY 1329 TGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGAGATCTTCTTTCTGCCATTCAATAATTC 1388  
Db 1125 TGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGAGATCTTCTTTCTGCCATTCAATAATTC 1184  
QY 1389 ATTGCATCATGGCATCCAGGCCAGCAATGATATGATCTACAAAAGGAGATC 1435  
Db 1185 ATTGCATCATGGCATCCAGGCCAGCAATGATATGATCTACAAAAGGAGGTC 1231  
RESULT 9  
AX664135  
LOCUS  
DEFINITION Sequence 7 from Patent WO02057304.  
ACCESSION AX664135  
VERSION AX664135.1 GI:29164145  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Panzer,S.R., Lincoln,S.E., Altus,C.M., Dufour,G.E., Jackson,J.L., Jones,A.L., Dam,T.C., Liu,T.F., Harris,B., Flores,V., Daffo,A., Marwaha,R., Chen,A.J., Chang,S.C., Gerstin,E.H., Peralta,C.H., David,M.H. and Lewis,S.A.  
TITLE Secretory molecules  
JOURNAL Patent: WO 02057304-A 7 25-JUL-2002;  
Incyte Genomics, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..719  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: LI:462889.1:2001JAN12"  
BASE COUNT 223 a 113 c 188 g 195 t  
ORIGIN  
Query Match 23.2%; Score 717.4; DB 6; Length 719;  
Best Local Similarity 99.9%; Pred. No. 9.9e-152;  
Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 474 GCGGGCGCGGAGCGCGTTCGTTATTTCCTGGTTCGGACAGTGGCGGCGCGGT 533  
Db 1 GCGGGCGCGGAGCGCGTTCGTTATTTCCTGGTTCGGACAGTGGCGGCGCGGT 60  
QY 534 GACCACGGGAGAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTACAGG 593  
Db 61 GACCACGGGAGAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTACAGG 120  
QY 594 ATGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTAC 653



FEATURES	source	94177 102536: contig of 8360 bp in length			
		102537 102636: gap of 100 bp			
		118954: contig of 16318 bp in length			
		119054: gap of 100 bp			
		147876: contig of 28822 bp in length.			
Location/Qualifiers					
misc_feature	1. .147876				
	/organism="Mus musculus"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:10090"				
	/clone="RP24-531M24"				
	/clone_lib="RPC1-24 Male Mouse BAC"				
	1. .84958				
	/note="assembly_fragment"				
	clone_end:SP6				
	vector_side:left"				
misc_feature	85059. .87411				
	/note="assembly_fragment"				
misc_feature	87512. .89933				
	/note="assembly_fragment"				
misc_feature	90034. .94076				
	/note="assembly_fragment"				
misc_feature	94177. .102536				
	/note="assembly_fragment"				
misc_feature	102637. .118954				
	/note="assembly_fragment"				
misc_feature	119055. .147876				
	/note="assembly_fragment"				
clone_end:T7					
vector_side:right"					
BASE COUNT		46111 a	27663 c	27986 g 45516 t 600 others	
ORIGIN					
Query Match 21.5%; Score 664.2; DB 2; Length 147876;					
Best Local Similarity 72.6%; Pred. No. 2.6e-139;					
Matches 1105; Conservative 0; Mismatches 293; Indels 125; Gaps 14;					
QY	1647	TGTTCTTTCAAGATCTCTGTTACAAACCAC-TTTCCTGGTGGATAACAAAGGTCCTTG	1705		
Db	72297	TGTTCTTTCAAGATCTCTGTTACAAACCACCTTTTCCTCGTGGTAAACAAAGATCTTTG	72356		
QY	1706	G-AACTCATCTCATGCAAGACATGGTGAAAGATGCACCTTCGGTCTTTTGTTCAGATCCTCC	1764		
Db	72357	GAAACTCCTCTCATGAAAGACATGGTGAAATATACCTTTGGTCTTCATT--AGTCCTC	72413		
QY	1765	GAGTGCCTTCCCCCAAGTGCTACCTATATAATAATCACCGGCTAAGGACTGTATCGACT	1824		
Db	72414	CAGTGCATCCCTTAAGTGTTCCTATATAATGATCACCGGCTAAGGACTGTATCGACT	72473		
QY	1825	CCTTTGTTACTCACTGTGTTTCGGCCATCTGTAGTCTTATTTCAGATCCATGGACATAACA	1884		
Db	72474	ACTTTGTTACTCACTGTGTTTGCCCATTTTGTAGTCTTGTTCAGAGTCATAGCCATAACA	72533		
QY	1885	GGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAATTTGCCACCTTGCAGGATG	1944		
Db	72534	GGGCTCAGCAGAGAGTCAAACCTTGGTCACATTCTTGAGAAAGTTTGTACCTTGCAAGATG	72593		
QY	1945	AGTTTATGAC-----ATTTTATTTTAATAGGCAGAGA	1976		
Db	72594	AGTTTATACTCCAGAGGATGAATGTTCTTGAAGAAATGTAAGAGTACACTTTGCAGACA	72653		
QY	1977	AGGTTGATGCAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATTTGG	2036		
Db	72654	ACTTTGATGCAGTGCTTCACACTCCGTTGTTGAAACAGGAGCCTGAGAGACTACATCTGA	72713		
QY	2037	CCTGTTTAGGTACCTGGGTCCTTTACCATAAACCTTCGCATTTATGATACAGTACCTTCTAA	2096		
Db	72714	CCTGTTTAGGAATCTGAGTCTCTTACCAGAAATCTTCAAATTATGATTCAGTATCTGCCCA	72773		
QY	2097	GTGGCTTTGAAATGGAACCTACAGTATGCACGAGTACTATTACATATATTGGTATCTCT	2156		
Db	72774	TTGCCCTTGAGCTGGAGCTGTACAGCAGCATGAGTTCTACTCCATGTACTGGTACCTCT	72833		

Qy	2157	CTGAATTCCTTTACGCATGGTTGATGTCAACATTGAGTCGTGCCGATGGCTCTCAATGG	2216	
Db	72834	CTGAGTTCTCTGATGCGTCGCTGATGTCAA-----ACCTGACAGCTCTCAGATGG	72883	
Qy	2217	CAGAGGAAAGGATAATGGAAGAGCAGCAGAGAAAGGCCGCTAGTAG-----	2259	
Db	72884	CAGAGGTGAAGATAATGGAGGAAACAGCAGAGAAAGGACACAGCAGAGAAACAAATAAACAA	72943	
Qy	2260	-----TAAAAAACAAGAAAAAAGAAAGTTCGCCCATTTGAGCCGAGAGATC	2308	
Db	72944	ACAAAAAACAACAAAAAAGAAAGAAAGTTCGTTGAGCCAAAGAGATC	73003	
Qy	2309	ACAATGAGCCAAAGCATATCAGAAACATGTGTCTGGAATGTTTAAACCATGGTAGCATTT	2368	
Db	73004	ATAATCAGCCAGGCCCTATCAGAA-ATGTGGCTGGAATGTCAGAAATGATGGTCATGTTT	73062	
Qy	2369	GACATGGACGGCAAGTACGTAAACCGAAGTTTGAGCTTGATAGTGAACAAGTTCGGTAT	2428	
Db	73063	GACATGGATGGCAAGTGCTCAAGCCCAAGTTTGAGCTTGATAGTATCAGGTCTGGTAT	73122	
Qy	2429	GAAACACAGGTTTGCTCCATTCAACAGTGTGATG-----ACCCGCGCCAGTGCAC	2480	
Db	73123	GAGCACAGGTCTGCTTTGTTCAACAGTGTGATGATGTGATTACACTGCCACAGTGAAC	73182	
Qy	2481	ACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATAATAGCCCTCCTCCTCAGTCTCCTG	2540	
Db	73183	ACCTGCCGTTCAAGGAAATGGCTGACCTCAATAAATAATAGCCCTTCTCTCAGGCTCCAG	73242	
Qy	2541	AACTGTATGTGGCAGCTAGTAAAGCACTTTCAACAGGCAAAATATGATTTGAAAAATATC	2600	
Db	73243	AGCTTTATGTGGCAGCTAATAAGCACTTTCAACAGGCAAAATATGATCCTTTGAAAAATG	73302	
Qy	2601	CTAACCCGGACCATGAGGTTTAAAGAAATTTTAAAGGTTGCCAAACCCAACTTTGTGTTA	2660	
Db	73303	CCAATCCAGGCAGTGAGGTCAGTAGAATCCTTAAGGTTGTTAAGTCCAACCTTTGTGCTTA	73362	
Qy	2661	TGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCTCCTGAATTTGATTTCT	2720	
Db	73363	TGAAGCTCTTGGCAGGAGGACACAAAAGAGAGTCAAAAGGTTCTCCCGAATTTGACTTCT	73422	
Qy	2721	CTGCTCATAAATATTTTCTGTGTGAAAACCTTGTGTTGAGAGAGACTGGGGAGGTGGCCAT	2780	
Db	73423	CTGTTCAAAAATACCTTTCATATGTTGAAAAGTGATTTGAAAGAGACTGCAAAAGACGACTGT	73482	
Qy	2781	AAAGGGCAGAGTCTTCTTTTCAGACCCCAACTCTTTAGAGGGCACATCACAGGCTCCACAT	2840	
Db	73483	GT-----CCACTCTAGGGGATACATCAGTCTGCCACAT	73515	
Qy	2841	CACGGGAAGTGAGATGGATTTCTTGGGT-----AACAACTCATTTAAGGAATACTTTT-A	2895	
Db	73516	ATGTTGAGGTAAAGCGGATTTCTTGAATGACAGACAGCTCCTTAAAAAAGAAATACTTTTCA	73575	
Qy	2896	GTTTGACAGCCTTATATGACATGAATGAAAACTGCTGTTTTAAAGTGGTTTATTATGTTT	2955	
Db	73576	GTTTAAACACCTTTACCTGGCATGATGAAAACTGCTCTCTTTAAAGTGGTTTATTATG---	73632	
Qy	2956	CATGGAAGAAACTGGTCTTTATTGAATGCATTGATGAACGTTTATATGTTTATTACAGAT	3015	
Db	73633	-----GGTGTAACCTGGGATACAAAGTATTGGTGAATGTTACATGATTTTATTATATAC	73685	
Qy	3016	TTAAT--CACAAATCATTTTTTATGAATGATTGAGTGAAATAGTGTTTATAAAGGTTAA	3073	
Db	73686	TTAATGTTTAAACACATTTTAAAGGAATGATTGAGTG-AAATGGTGTGTTGTTGAAAGATTAA	73744	
Qy	3074	TAAATTTCTTGACAAAAAATAA 3096		
Db	73745	TAAATTTCTTGACAAAAAATAA 73767		

RESULT 11  
AL353743  
LOCUS Human DNA sequence from clone RP11-213G2 on chromosome 9, complete  
DEFINITION

```
sequence.
AL353743
AL353743.22 GI:15552931
HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 181864)
Corby,N.
Direct Submission
Submitted (08-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Sep 11, 2001 this sequence version replaced gi:14267784.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-213G2 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-213G2 It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone RP11-213G2 is at 181864 in this
sequence. The true left end of clone RP11-65C15 is at 111320 in
this sequence. The true right end of clone RP11-202111 is at 100 in
this sequence.
Location/Qualifiers
1. 181864
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-213G2"
/clone_lib="RPCI-11.1"
39727
unsure
misc_feature 115957..116075
/note="Sequence from overlapping clone BA65C15 (AL161453).
Assembly confirmed by restriction digest."
misc_feature 117488..117564
/note="Sequence from overlapping clone BA65C15 (AL161453).
Assembly confirmed by restriction digest."
BASE COUNT 53151 a 36268 c 36336 g 56109 t
ORIGIN
Query Match 17.8%; Score 551.4; DB 9; Length 181864;
Best Local Similarity 99.8%; Pred. No. 1.1e-113;
Matches 552; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TTCCTCAGGAACTCCAGCGCTGTATAGGAAACATAAATCCGTGTCTCAGGCAGCAGTA 60
|||||
Db 174760 TTCCTCAGGAACTCCAGCGCTGTATAGGAAACATAAATCCGTGTCTCAGGCAGCAGTA 174819
```

```
Qy 61 GCACGCTGTTGCTCTCGGAGCTTGGCTGCTCGTTCGTCTCGCAACCACTAAGGTCTACG 120
|||||
Db 174820 GCACGCTGTTGCTCTCGGAGCTTGGCTGCTCGTTCGTCTCGCAACCACTAAGGTCTACG 174879
Qy 121 CAAACCTCCACGGTTTCTTCCGCTTCGCGTTCACCTTTCTAAGAAATCCAGAGGGCA 180
|||||
Db 174880 CAAACCTCCACGGTTTCTTCCGCTTCGCGTTCACCTTTCTAAGAAATCCAGAGGGCA 174939
Qy 181 GCGCAGACGGGGGGCTCTGAGACTCCGGGGCTCCGCCTCTTTCCGGAAACCGCCCACTA 240
|||||
Db 174940 GCGCAGACGGGGGGCTCTGAGACTCCGGGGCTCCGCCTCTTTCCGGAAACCGCCCACTA 174999
Qy 241 CCCAGGACTCCGACAGAGGGTGAAAAAAGATAACTTCCGGTCTCGCATCGTCTCTAATC 300
|||||
Db 175000 CCCAGGACTCCGACAGAGGGTGAAAAAAGATAACTTCCGGTCTCGCATCGTCTCTAATC 175059
Qy 301 TCGCGAGAAGAGAGCGCGCCGCTCCCGGAGTAAGGAGGCGGTGGCGAGGGAGGGGGTG 360
|||||
Db 175060 TCGCGAGAAGAGAGCGCGCCGCTCCCGGAGTAAGGAGGCGGTGGCGAGGGAGGGGGTG 175119
Qy 361 TGGCCGGGAGCGCGAAGTCCCGGAGTAAGGAGGAGGGGCGGGTCCGCTCCCGG 420
|||||
Db 175120 TGGCCGGGAGCGCGAAGTCCCGGAGTAAGGAGGAGGGGCGGGTCCGCTCCCGG 175179
Qy 421 GCATACGCATGCGTGACACGCTGCCGCTCGGGCTGGGCTGAGAGGGAGGGGCGCGCG 480
|||||
Db 175180 GCATACGCATGCGTGACACGCTGCCGCTCGGGCTGGGCTGAGAGGGAGGGGCGCGCG 175239
Qy 481 GCGGAGCGCGCTCGTTATTTCCGTGGTCCGGACAGTGGTGGCGCGCGGTGACACAG 540
|||||
Db 175240 GCGGAGCGCGCTCGTTATTTCCGTGGTCCGGACAGTGGTGGCGCGCGGTGACACAG 175299
Qy 541 GGAGAAAGTAGGCA 553
Db 175300 GGAGAAAGTAGGTA 175312
```

```
RESULT 12
AL161447
LOCUS AL161447 160115 bp DNA linear PRI 16-MAR-2002
DEFINITION Human DNA sequence from clone RP11-379P1 on chromosome 9, complete
sequence.
ACCESSION AL161447
VERSION AL161447.20 GI:19572746
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 160115)
Johnston,C.
Direct Submission
Submitted (16-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Mar 21, 2002 this sequence version replaced gi:15787718.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
```

```
COMMENT
```



ACCESSION AJ334413  
 VERSION AJ334413.1 GI:15878831  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 743)  
 AUTHORS Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F., Kvasha, S.M.,  
 Podowski, R.M., Matushkin, Y.G., Gyanchandani, A., Muravenko, O.V.,  
 Levitsky, V.G., Kolchanov, N.A., Protopopov, A.I., Kashuba, V.I.,  
 Kisselev, L.L., Wasserman, W., Wahlestedt, C. and Zabarovsky, E.R.  
 TITLE NotI flanking sequences: a tool for gene discovery and verification  
 of the human genome  
 JOURNAL Nucleic Acids Res. 30 (14), 3163-3170 (2002)  
 MEDLINE 22131767  
 PUBMED 12136098  
 REFERENCE 2 (bases 1 to 743)  
 AUTHORS Zabarovsky, E.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorigenesis Centre,  
 Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,  
 Sweden  
 FEATURES  
 source Location/Qualifiers  
 1. .743  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="NR1-ND24R"  
 BASE COUNT 219 a 154 c 166 g 203 t 1 others  
 ORIGIN  
 Query Match 9.7%; Score 299; DB 9; Length 743;  
 Best Local Similarity 95.4%; Pred. No. 7.9e-57;  
 Matches 308; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
 QY 1 TTCCTCAGAACTCCAGCGCTGTATAGAAACATAAAATCCGTTGTCAGGACGAGTA 60  
 Db |||||  
 323 TTCCTCAGGAACTCCAGCGCTGTATAGAAACATAAAATCCGTTGTCAGGACGAGTA 264  
 QY 61 GCACGCTGTTGCTCTCGGAGCTTGGCTGCTCGTTCGTTCGCAACCACTAAGTCTACG 120  
 Db |||||  
 263 GCACGCTGTTGCTCTCGAAGCTTGGCTGTTCCGTCGTTCGCAACCACTAAGTCTACG 204  
 QY 121 CAAACCTCCACGGTTTCCTTCCGCCCTTCGCGTCACTTCTAAGAAATCCGAGGGCA 180  
 Db |||||  
 203 CAAACCTCCACGGTTTCCTTCCGCCCTTCGCGTCACTTCTAAGAAATCCGAGGGCA 144  
 QY 181 GCGCAGACGGGGGGGCTCTGAGACTCCGGGCTCCGCCCTCTTTCGGGAACCGCCCACTA 240  
 Db |||||  
 143 GCGCAGACGGGGGGGCTCTGAGACTCCGGGCTCCGCCCTCTTTCGGGAACCGCCCACTA 84  
 QY 241 CCCAGGACTCCGACAGAGGGTGAAAAAGATAACTTCGGGTCTCGCGATCGTCTTAATC 300  
 Db |||||  
 83 CTCAGGACTCCGTCAGAGAAAAAACCGATTACTTCGGGTCTCGCGATCGTCTTAATC 24  
 QY 301 TCGCGAAGAGAGGCGGCCGC 323  
 Db |||||  
 23 TCGCGAAGAGAGGCGGCCGC 1

```

REFERENCE
AUTHORS
Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
Levitsky,V.G., Kolchanov,N.A., Protopopov,A.I., Kashuba,V.I.,
Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
TITLE
NotI flanking sequences: a tool for gene discovery and verification
of the human genome
JOURNAL
Nucleic Acids Res. 30 (14), 3163-3170 (2002)
MEDLINE
22131767
PUBMED
12136098
REFERENCE
AUTHORS
Zabarovsky,E.R.
TITLE
Direct Submission
JOURNAL
Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
Sweden
FEATURES
source
Location/Qualifiers
1..642
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="NR1-ND24C"
BASE COUNT
73 a 209 c 265 g 93 t 2 others
ORIGIN
Query Match 7.6%; Score 236.4; DB 9; Length 642;
Best Local Similarity 99.6%; Pred.No. 1.2e-42;
Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 316 GCGGCGCCATCGGCCGAAACGGAGCGGTGGCGAGGGAGGGGGTGTGCGCCGGGAGCGCG 375
Db 1 GCGGCGCCATCGGCCGAAACGGAGCGGTGGCGAGGGAGGGGGTGTGCGCCGGGAGCGCG 60
QY 376 AAGTCCCGGGAGTAAGGGAGAGGGGCGGGGTGCGCGCTCCGGGCATACGCATGCGTG 435
Db 61 AAGTCCCGGGAGTAAGGGAGAGGGGCGGGGTGCGCGCTCCGGGCATACGCATGCGTG 120
QY 436 CACGTGCGCGTGGGCTGGGCTGAGAGGGAGGGGGCGGGCGCGCGCGGCGGCGTGG 495
Db 121 CACGTGCGCGTGGGCTGGGCTGAGAGGGAGGGGGCGGGCGCGCGCGGCGGCGTGG 180
QY 496 TTATTTCGTTGGTCCGGACAGTGGTGGCGCGCGGGGTGACCAACGGGAGAGTAGGCA 553
Db 181 TTATTTCGTTGGTCCGGACAGTGGTGGCGCGCGGGGTGACCAACGGGAGAGTAGGTA 238

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 25, 2003, 03:13:51 ; Search time 106 Seconds  
(without alignments)  
10683.169 Million cell updates/sec

Title: US-10-001-857-42  
Perfect score: 5439  
Sequence: 1 ttcttcacgaactccagg.....atttcttgacaaaaaaa 3096

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 667484 seqs, 182883703 residues

Total number of hits satisfying chosen parameters: 1334968

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US10001857/runat\_24112003\_162856\_9623/app\_query\_fasta\_1.3271  
-DB=Published Applications\_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosu62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10001857@cgn\_1\_149@runat\_24112003\_162856\_9623  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	3796	69.8	733	14	US-10-001-857-145	Sequence 145, Appl
2	165	3.0	523	12	US-10-017-161-1982	Sequence 1982, Ap
3	146.5	2.7	574	12	US-10-168-097A-76	Sequence 76, Appl
4	146.5	2.7	574	12	US-10-239-431A-38	Sequence 38, Appl
5	145	2.6	559	10	US-09-858-155A-2	Sequence 2, Appl
6	145	2.6	559	12	US-10-168-097A-26	Sequence 26, Appl
7	145	2.6	559	12	US-10-239-431A-33	Sequence 33, Appl
8	145	2.6	559	12	US-10-116-370-2	Sequence 2, Appl
9	144	2.6	1134	14	US-10-001-873-50	Sequence 50, Appl
10	142.5	2.6	800	12	US-10-029-386-32198	Sequence 32198, A
11	142	2.6	721	14	US-10-086-464-5	Sequence 5, Appl
12	141.5	2.6	5877	15	US-10-142-515-11	Sequence 11, Appl
13	141.5	2.6	5935	15	US-10-243-243A-8	Sequence 8, Appl
14	140	2.5	802	9	US-09-823-240-2	Sequence 2, Appl
15	139.5	2.5	1243	15	US-10-196-935A-4	Sequence 4, Appl
16	138.5	2.5	641	11	US-09-959-987-2	Sequence 2, Appl
17	138.5	2.5	647	9	US-09-925-299-1002	Sequence 1002, Ap
18	138.5	2.5	647	11	US-09-925-299-1002	Sequence 2, Appl
19	136.5	2.5	647	14	US-10-086-464-2	Sequence 4, Appl
20	136.5	2.5	647	14	US-10-086-464-4	Sequence 36, Appl
21	134.5	2.4	520	12	US-10-168-097A-36	Sequence 34, Appl
22	134.5	2.4	520	12	US-10-239-431A-34	Sequence 22, Appl
23	134	2.4	351	12	US-10-239-431A-22	Sequence 10, Appl
24	134	2.4	542	12	US-10-239-431A-10	Sequence 24, Appl
25	133.5	2.4	507	14	US-10-078-547-24	Sequence 14830, A
26	132.5	2.4	206	15	US-10-156-761-14830	Sequence 10907, A
27	132.5	2.4	797	15	US-10-156-761-10907	Sequence 2, Appl
28	132	2.4	503	14	US-10-078-547-2	Sequence 20, Appl
29	131	2.4	528	12	US-09-840-746-20	Sequence 43, Appl
30	131	2.4	538	10	US-09-976-740-43	Sequence 43, Appl
31	131	2.4	538	14	US-10-023-529-43	Sequence 43, Appl
32	131	2.4	538	14	US-10-023-523-43	Sequence 102, Appl
33	130	2.4	235	15	US-10-050-704-102	Sequence 7399, Ap
34	130	2.4	455	12	US-10-032-585-7399	Sequence 46, Appl
35	129.5	2.4	501	12	US-10-168-097A-46	Sequence 35, Appl
36	129.5	2.4	501	12	US-10-239-431A-35	Sequence 7, Appl
37	129	2.3	377	15	US-10-149-819-7	Sequence 56, Appl
38	129	2.3	505	12	US-10-168-097A-56	Sequence 36, Appl
39	129	2.3	505	12	US-10-239-431A-36	Sequence 2256, Ap
40	126.5	2.3	405	12	US-10-017-161-2256	Sequence 10088, A
41	126.5	2.3	1046	15	US-10-156-761-10088	Sequence 282, Appl
42	126	2.3	416	15	US-10-043-487-282	Sequence 12, Appl
43	126	2.3	505	12	US-10-168-097A-12	Sequence 32, Appl
44	126	2.3	505	12	US-10-239-431A-32	Sequence 47, Appl
45	125	2.3	550	10	US-09-976-740-47	

ALIGNMENTS

RESULT 1  
US-10-001-857-145  
; Sequence 145, Application US/10001857  
; Publication No. US20020183500A1  
; GENERAL INFORMATION:  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herve  
; APPLICANT: Chen, Sei-Yu  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P  
; FILE REFERENCE: DEX-0273  
; CURRENT APPLICATION NUMBER: US/10/001,857  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/252,054  
; PRIOR FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 208  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 145  
; LENGTH: 733  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-001-857-145

Alignment Scores:				
Pred. No.:	5e-294	Length:	733	
Score:	3796.00	Matches:	732	
Percent Similarity:	99.59%	Conservative:	0	
Best Local Similarity:	99.59%	Mismatches:	1	
Query Match:	69.79%	Indels:	3	
DB:	14	Gaps:	0	
US-10-001-857-42 (1-3096) x US-10-001-857-145 (1-733)				
QY	556	ATGGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGGATGGGAGCTCAGTATGCCAGAA	615	
Db	1	MetValMetLysAlaSerValAspAspAspSerGlyTrpGluLeuSerMetProGlu	20	
QY	616	AAAAATGAGAAAAAGCAATACAAACTGGGTGGACATTACCCCAAGATTTTGAAGAAGCTTGT	675	
Db	21	LysMetGluLysSerAsnThrAsnTrpValAspIleThrGlnAspPheGluGluAlaCys	40	
QY	676	CGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCTATTTTGGTCTTTTGAAGCCATG	735	
Db	41	ArgGluLeuLysLeuGlyGluLeuLeuHisAspLysLeuPheGlyLeuPheGluAlaMet	60	
QY	736	TCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAACCAAGTTAAT	795	
Db	61	SerAlaIleGluMetMetAspProLysMetAspAlaGlyMetIleGlyAsnGlnValAsn	80	
QY	796	CGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCAGCTATTAAAAATTAAAGATCTC	855	
Db	81	ArgLysValLeuAsnPheGluGlnAlaIleLysAspGlyThrIleLysIleLysAspLeu	100	
QY	856	ACCTTGCCGTGAAGCTGATAGGGATTATGGATACATGTTTTTGGCTGTTTGATAACGTGGTTA	915	
Db	101	ThrLeuProGluLeuIleGlyIleMetAspThrCysPheCysLeuIleThrTrpLeu	120	
QY	916	GAAGGCCATTCACTGGCACAGACAGATATTACGTGCCTTTACATTCAATAATCCAGACTTT	975	
Db	121	GluGlyHisSerLeuAlaGlnThrValPheThrCysLeuTyriIleHisAsnProAspPhe	140	
QY	976	ATAGAAGATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAAATCTGTGACATTGCA	1035	
Db	141	IleGluAspProAlaMetLysAlaPheAlaLeuGlyIleLeuLysIleCysAspIleAla	160	
QY	1036	AGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTCAATGACTTAT	1095	
Db	161	ArgGluLysValAsnLysAlaAlaValPheGluGluGluAspPheGlnSerMetThrTyr	180	
QY	1096	GGATTTAAATCGCTAACAGTGTGACAGATCTTCGAGTTACAGGCATGCTAAAAGATGTG	1155	
Db	181	GlyPheLysMetAlaAsnSerValThrAspLeuArgValThrGlyMetLeuLysAspVal	200	
QY	1156	GAGGATGACATGCAAAAGAGAGTAAAGAGTACTCGAAGTCGACAAGGAGAAAGAGAT	1215	
Db	201	GluAspAspMetGlnArgArgValLysSerThrArgSerArgGlnGlyGluGluArgAsp	220	
QY	1216	CCAGAAGTTGAAGTGAACACCAACAATGTTTAGCAGTATTCAGCAGAGTGAAATTTACT	1275	
Db	221	ProGluValGluLeuGluHisGlnGlnCysLeuAlaValPheSerArgValLysPheThr	240	
QY	1276	CGTGTGTTACTGACAGTGTCTTATAGCCTTTACTAAGAAAGAGACCAGTGCTGTGCAGAA	1335	
Db	241	ArgValLeuLeuThrValLeuIleAlaPheThrLysLysLysGluThrSerAlaValAlaGlu	260	
QY	1336	GCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATTCATAATTCATTGCAT	1395	
Db	261	AlaGlnLysLeuMetValGlnAlaAlaAspLeuLeuSerAlaIleHisAsnSerIleuHis	280	
QY	1396	CATGGCATCCAGGCCCGCAGAAATGATACTACAAAAGGAGATCATCCAATTATGATGGTTTT	1455	
Db	281	HisGlyIleGlnAlaGlnAsnAspThrThrLysGlyAspHisProIleMetMetGlyPhe	300	
QY	1456	GAACCCCTTGTGAACAGAGGCTACTTCCACCTACCTTCCCTCGATATGCAAAAATAATT	1515	
Db	301	GluProLeuValAsnGlnArgLeuLeuProProThrPheProArgTyrAlaLysIleIle	320	

QY	1516	AAAAGGGAAGAAATGGTGAACCTATTTTGCAGATTAAATAGATAGAAATAAAAACTGTCTGT	1575	
Db	321	LysArgGluGluMetValAsnTyrPheAlaArgLeuIleAspArgIleLysThrValCys	340	
QY	1576	GAGGTTGTGAATTTAACAAAATTTACATTGTATCCTGGATTTTCTGTGAATTTTAGTGAA	1635	
Db	341	GluValValAsnLeuThrAsnLeuHisCysIleLeuAspPhePheCysGluPheSerGlu	360	
QY	1636	CAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACCTTTCCTGGTGATAACAAA	1695	
Db	361	GlnSerProCysValLeuSerArgSerLeuLeuGlnThrThrPheLeuValAspAsnLys	380	
QY	1696	AAGGTTCTTTGGAACCTCATCTCATGCAAGACATGGTGAAAGATGCACCTTCGGTCTTTTGT	1755	
Db	381	LysValPheGlyThrHisLeuMetGlnAspMetValLysAspAlaLeuArgSerPheVal	400	
QY	1756	AGATCCTCCGAGTGCTTTCCCCCAAGTGCTACCTATATATAATCAACACGAGCTAAGGACT	1815	
Db	401	-AspProPro-ValLeuSerProLysCysTyrLeuTyrAsnAsnHisGlnAlaLysAspC	420	
QY	1816	GTATCGACTCCTTTGTACTCACTGTGTTCCGCCATTCTGTAGTCTTATTTCAGATCCATG	1875	
Db	420	ysIleAspSerPheValThrHisCysValArgProPheCysSerLeuIleGlnIleHisG	440	
QY	1876	GACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAATTTGCCACCT	1935	
Db	440	lyHisAsnArgAlaArgGlnArgAspLysLeuGlyHisIleLeuGluGluPheAlaThrL	460	
QY	1936	TGCAGGATGAGTTTATGACATTTTATTTAATAGGCA - GAGAAAGTTGATGCAGCGCTTC	1994	
Db	460	euGlnAspGluPheMetThrPheTyrPheAsnArgAlaGluLysValAspAlaAlaLeuH	480	
QY	1995	ACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATTTTGGCCCTGTTTAGGTACCTGGG	2054	
Db	480	isThrMetLeuLeuLysGlnGluProGlnArgGlnHisLeuAlaCysLeuGlyThrTrpV	500	
QY	2055	TCCTTTACATAAACCTTCGCATTATGATACAGTACCTTCTAAGTGGCTTTTGAATTTGGAAC	2114	
Db	500	alLeuTyrHisAsnLeuArgIleMetIleGlnTyrLeuLeuSerGlyPheGluLeuGluL	520	
QY	2115	TCTACAGTATGCACGAGTACTATTACATATATTGGTATCTCTCTGAATTCCTTTACGCAT	2174	
Db	520	euTyrSerMetHisGluTyrTyrTyrIleTyrTrpTyrLeuSerGluPheLeuTyrAlat	540	
QY	2175	GGTTGATGTCAACATTGAGTCGTGCGCATGGCTCTCAAAATGGCAGAGAAAGGATAATGG	2234	
Db	540	rpLeuMetSerThrLeuSerArgAlaAspGlySerGlnMetAlaGluGluArgIleMetG	560	
QY	2235	AAGAGCAGCAGAAAAGCCGTAGTAGTAAAAAACAAGAAAAAAGAAAGTTTCGCCCAT	2294	
Db	560	luGluGlnGlnLysGlyArgSerSerLysLysThrLysLysLysLysValArgProL	580	
QY	2295	TGAGCCGAGAGATCACAATGAGCCCAAGCATATCAGAACATGTGTGCTGGAAATGTTTAAA	2354	
Db	580	euSerArgGluIleThrMetSerGlnAlaTyrGlnAsnMetCysAlaGlyMetPheLyst	600	
QY	2355	CCATGGTAGCATTTGACATGGACGGCAAGTACGTAAACCGAAGTTTGAGCTTGATAGTG	2414	
Db	600	hrMetValAlaPheAspMetAspGlyLysValArgLysProLysPheGluLeuAspSerG	620	
QY	2415	AACAAGTTCGGTATGAACACAGGTTTGTCTCCATTCAACAGTGTGATGACCCCGCCGAG	2474	
Db	620	luGlnValArgTyrGluHisArgPheAlaProPheAsnSerValMetThrProProProv	640	
QY	2475	TGCACCTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTCCTCAGT	2534	
Db	640	alHisTyrLeuGlnPheLysGluMetSerAspLeuAsnLysTyrSerProProGlns	660	
QY	2535	CTCCTGAACTGTATGTGGCAGCTAGTAAGCACCTTTCAACAGGCAAAATATATTGGAAA	2594	
Db	660	erProGluLeuTyrValAlaAlaSerLysHisPheGlnGlnAlaLysMetIleLeuGluA	680	







```
QY 454 CAGCCCGACCGGACGGTGCACGCATGCGTATGCCGGGACGGCGGACCC-----GCC 401
Db 334 AlaLeuSerThrSerSerLeuAlaSerMetThrSerThrProProProValPro 353

QY 400 CCCTCTCCCTTACTCCGGGACTTCG---CGCTCCCGGGCCACACCCCT----- 353
Db 354 ProProProProProAlaThrAlaLeuGlnAlaProAlaValProProProAla 373

QY 352 -----CCCTCGCCACCG-----CCTCCGTTTC 332
Db 374 ProLeuGlnIleAlaProGlyValLeuHisProAlaProProProIleAlaProLeu 393

QY 331 GGCCGATGGCGCGCCTTCTCTCTCGCGAGATTAGAGACGATCGCGAGACCGGAAGTT 272
Db 394 ValGlnProSerProPro-----ValAlaArgAlaAlaProValCys-----GluThr 409

QY 271 ATCTTTTTTCCACCTCTGTGCGAGTCCTGGGTAGTGGCGGTTCCTCCGGAAGAGCGGAG 212
Db 410 ValProValHisProLeuProGlnGlyGluValGlnGlyLeuProProProProPro 429

QY 211 CCCGGAGTCTCAGAGCCCGCCCGTCTGCGCTGCCCTCT 173
Db 430 ProProLeu-----ProProProGlyIleArgProSer 440

RESULT 8
US-10-116-370-2
; Sequence 2, Application US/10116370
; Publication No. US20030190709A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: PABLO, A POLYPEPTIDE THAT INTERACTS WITH BCL-XL, AND
; TITLE OF INVENTION: USES RELATED THERETO
; FILE REFERENCE: AM100012-D2
; CURRENT APPLICATION NUMBER: US/10/116,370
; CURRENT FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 559
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-370-2

Alignment Scores:
Pred. No.: 0.00942 Length: 559
Score: 145.00 Matches: 48
Percent Similarity: 40.52% Conservative: 14
Best Local Similarity: 31.37% Mismatches: 61
Query Match: 2.64% Indels: 30
DB: 12 Gaps: 8

US-10-001-857-42 (1-3096) x US-10-116-370-2 (1-559)

QY 559 CCATTATGCCTA-----CTTCTCCCGTGGTCAACCCGCGCGCCACGCACT 515
Db 294 ProThrCysIleSerSerAlaThrGlyLeuIleGluAsnArgProGlnSerProAlaThr 313

QY 514 GTCCGGACACGGAAATAACGACGCCCTCGGCGCGCGCCCTCCCTCCCTCTCAGCC 455
Db 314 GlyArgThrProValPheValSerProThrProProProProProProLeuProSer 333

QY 454 CAGCCCGACCGGCAGCGTGACGCATGCGTATGCCGGGACGGCGGACCC-----GCC 401
Db 334 AlaLeuSerThrSerSerLeuAlaSerMetThrSerThrProProProProValPro 353

QY 400 CCCTCTCCCTTACTCCGGGACTTCG---CGCTCCCGGGCCACACCCCT----- 353
Db 354 ProProProProProAlaThrAlaLeuGlnAlaProAlaValProProProAla 373

QY 352 -----CCCTCGCCACCG-----CCTCCGTTTC 332
Db 374 ProLeuGlnIleAlaProGlyValLeuHisProAlaProProProIleAlaProLeu 393
```

```
QY 331 GGCCGATGGCGCGCCTTCTCTTCTCGCGAGATTAGAGACGATCGCGAGACCGGAAGTT 272
Db 394 ValGlnProSerProPro-----ValAlaArgAlaAlaProValCys-----GluThr 409

QY 271 ATCTTTTTTACCCCTGTGCGAGTCCTGGGTAGTGGCGGTTCCTCCGGAAGAGCGGAG 212
Db 410 ValProValHisProLeuProGlnGlyGluValGlnGlyLeuProProProProPro 429

QY 211 CCCGGAGTCTCAGAGCCCGCCCGTCTGCGCTGCCCTCT 173
Db 430 ProProLeu-----ProProProGlyIleArgProSer 440

RESULT 9
US-10-001-873-50
; Sequence 50, Application US/10001873
; Publication No. US20020160388A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P:
; FILE REFERENCE: DEX-0275
; CURRENT APPLICATION NUMBER: US/10/001,873
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,055
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,496
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 1134
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-873-50
```

```
Alignment Scores:
Pred. No.: 0.0145 Length: 1134
Score: 144.00 Matches: 53
Percent Similarity: 37.97% Conservative: 7
Best Local Similarity: 33.54% Mismatches: 64
Query Match: 2.62% Indels: 34
DB: 14 Gaps: 8
```

US-10-001-857-42 (1-3096) x US-10-001-873-50 (1-1134)

```
QY 559 CCATTATGCCTACTTCTCCCGTGGTCAACCCGCGCGCCACGCACTGTCCGGACCGGAA 500
Db 976 ProLeuProLeuArgLeuProProProProProProPro----- 988

QY 499 ATAAAGACGGCGCTCGGCGCGCGCGCCCTCCCTCTCTCA-----GCCAGCCCGAC 446
Db 989 ---LeuProArgProHisProProProProProProProProLeuLeuProProPro 1007

QY 445 CGGCAGCGTGCA-----CGCATGCGTATGCCCGGACGCGCGACCCCGCC 401
Db 1008 GlnThrArgThrLeuProAlaAlaArgThrMetArgGlnPro-----ProPro 1023

QY 400 CCCTCTCCCTTACTCCCGGGACTTCGCGCTCCCGGCCACACCCCTCCCTCGCCA--- 344
Db 1024 ProArgLeuAlaLeuProArgArgArgArgSerProProArgProProSerArgProAla 1043

QY 343 -----CCGCCTCCGTTCCGCGCGGATGGCGCGCCTTCTCTTCTCGCGAGA 299
Db 1044 ArgArgGlyProArgProThrProGlnAlaArgArgArgProArgProSerProArgArg 1063

QY 298 TTAGAGACGATCGCGAGACCGGAAGTTATCTTTTTCACCTCTGTTCGGAGTCTCGGTA 239
Db 1064 Leu-----LeuArgSerPro-----HisSerLeuCysSerProArgLeu 1076
```

```
QY 238 GTGGCGGTTCCGGAAAGAG-----GCGGAGCCCGGAGTCTCAGAGCCCGCCCGCTCT 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1077 ArgProGlyProArgAlaAspProArgArgGluArgAlaSerThrSerProProProArg 1096
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 GCGTCCCTCTGGGAATTTCTAGAAAGGTGACGCGAAGGCGGAAGGAAACCG 131
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1097 SerTrpProSerGlySerAlaCysArgProTrpArgThrGlyProArgSerPro 1114
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-10-029-386-32198
; Sequence 32198, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32198
; LENGTH: 800
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007663.28
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
US-10-029-386-32198

Alignment Scores:
Pred. No.: 0.0169 Length: 800
Score: 142.50 Matches: 76
Percent Similarity: 35.66% Conservative: 77
Best Local Similarity: 17.72% Mismatches: 202
Query Match: 2.59% Indels: 74
DB: 12 Gaps: 11

US-10-001-857-42 (1-3096) x US-10-029-386-32198 (1-800)

QY 1334 TCTGCAACAGCACTGGTCTCTTCTTAGTAAGGCTATAAGCACTGTCAGTAACACACGA 1275
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 290 ThrThrThrIleThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 309
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1274 GTAAATTTC-----ACTCTGCTGAATACTGCTAAACATTGTTGGTGTCTAGTTCA 1224
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 IleThrIleThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 329
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1223 ACTTCTGGATCTCTTTCTCTCTTGTGCGACTTCGAGTACTCTTTTACTCTTCTTTGCA 1164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330 ThrThrThrIleThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1163 TCATCCTCCACATCTTTTAGCATGCTGTAACTCGAAGATCTGTACACTGTAGCCATT 1104
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 350 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 369
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1103 TTAAATCCATAAGTCATGTAAGTAAATCTCTTCAAAAACAGCAGCTTTATTACT 1044
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 IleThrThr-----ThrThrThrIleThrThrThrThrThrThrThrThrThrThr 384
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1043 TTTTCCCTTGCAATGTCACAGATTTTCAAGATTCCAGAGCAAAAGCCTTATAGCAGGA 984
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 404
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 983 TCTTCTATAAAGCTGGATTATGAATGTAAGGCAGCGTAATAACTGTCTGTGCCAGTGAA 924
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 405 ThrThrIleThrThrThrIle-----ThrThrThrThrIleThrThrThrThr 419
QY 923 TGGCCTTCTAACCACGTTATCAAAACAGCAAAAACATGTATCCATAATCCCTATCAGTTCA 864
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 ThrThrThrThrIleThrSerThrThrThrThrThrThrThrThrThrThrThrThr 439
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 863 GGCAAGGTGAGATCTTTAATTTTAATAGTGCCATCCTTGATAGCTTGTTCAAAATTGAGA 804
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 440 -----ThrThrThrIleThrIleThrSerThrThrThrThrThrThrThrIleThr 455
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 803 ACTTTTCGATTAACTTGGTTTCCAATCATGCCAGCATCCATCTTGGGATCCATCTTCA 744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 ThrThrThrThrThrThrIleThrIleThrSerThrThr-----ThrThrThrThr 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 743 ATAGCAGACATGGCTTCAAAAAGACCATAATAGCTTATCATGAAGTAGTTCTCCCAACTTT 684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 473 IleThrThrThrThrThrArgThrThrIleThrThrThrThrThrThrThrThrThr 492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 683 AATCTCGACAAGCTTCTTCAAAATCTTGGGTAATGTCCACCAGTTTGTATGCTTTTC 624
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 493 ThrThrIleThrThrThrThrIleThrIleThrThrThrThrThrThrThrThrThr 512
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 623 TCCATTTTCTGGCATACTGAGCTCCCATCCTGAATCGTCATCATCTACAGAAGCTTTC 564
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 513 ThrThrThrIleThrIleThrThrThrThrThrThrThrThrThrThrThrThrThr 532
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 563 ATAACCATTTATGCCTACTTCTCCCGTGGTCACCCCGCGCCGCCAGCTGTCCGGACCAC 504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 533 IleThrIleThrThrThrThrThrIleThrThrThrThrThrThrThrThrThrThr 552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 503 G-----GAAATAACGACGCGCCT-----CG 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 552 rThrThrIleThrIleThrSerThrThrThrThrThrAlaIleThrThrThrThrIle 572
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 502 -----GAAATAACGACGCGCCT-----CG 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 572 eThrIleThrThrThrProProLeuProProSerProProProProProSerLeuPr 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 482 GCGCGCGCGCCCTCCCTCTCAGCCCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 592 oProProProProProSerLeuProProProProProProProProProProProLe 612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 422 GCGCGGACGCGCGACCCCGCCCCCTCTCCCTTACTC-----CCGGGGAC 378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 612 uProSerLeuProProProProProProProProProProProProProProProSerPr 632
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 377 TTCGCGCTCCCGGCGCACACCCCTCCCTCG---CCACCGCCTCCGTTCCGCGCGATGGCG 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 632 oSerLeuProProProProProProProProProProProProProProProSerLeuPr 652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 320 GCGGCTTCTCTTCTCGGAGATTAGAGACGATCGGAGACCGGAAAGTTATCTTTTTTCA 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 652 oProProSerLeuProProSerProProSerProProSerLeuSerProProSerLeu----- 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 260 CCCTCTGTGCGAGTCTGGGTAGTGGCGGTTCCCGGAAAGAGCGGAGCCCGGAGTCTC 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 669 -ProLeu-----ProProProProProSerLeuPr 678
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 200 AGAGCCCGCGCGTCTGCGCTGCCC 176
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 678 oProProProProSerLeuProPro 686
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 11
US-10-086-464-5
; Sequence 5, Application US/10086464
; Publication No. US20020199218A1
; GENERAL INFORMATION:
; APPLICANT: GORING, Daphne R. et al.
; TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES
; FILE REFERENCE: P 25,762-A USA
; CURRENT APPLICATION NUMBER: US/10/086,464
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 10/069,304
```

```
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: PCT/CA00/00966
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: US 60/149,466
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: US 60/159,122
; PRIOR FILING DATE: 1999-10-13
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 721
; TYPE: PRT
; ORGANISM: Brassica napus
US-10-086-464-5

Alignment Scores:
Pred. No.:      0.0178      Length:      721
Score:          142.00      Matches:      58
Percent Similarity: 37.04%      Conservative: 22
Best Local Similarity: 26.85%      Mismatches: 75
Query Match:    2.58%      Indels:      61
DB:              14      Gaps:      10

US-10-001-857-42 (1-3096) x US-10-086-464-5 (1-721)
QY 788 TGGTTTCCAATCATGCCAGCATCCATCCTTTGGGATCCATCATTTCAATAGCAGACATGGCT 729
Db 4 TtpSerProCysLeuLeuSerSerProAlaAlaSerPheHisSerLeu----- 19
QY 728 TCAGAAAAGACCAATAGCTTATCATGAAGTAGTTCTCCCACTTTAATTCTCGACAAGCT 669
Db 20 -----HisLeuProPhePheArgSerLysGluAla 29
QY 668 TCTTCAAAATCTTGGGTAATGTCACCCAGTTTGTATTGCTTTTCTCCATTTTTCCTGGC 609
Db 30 AspMetSerSer-----AlaProSerProGlyThrGly 40
QY 608 ATACTGAGTCCCATCCTGAATCGTCATCATCTACA-----GAAGCTTTCATAACCAATT 555
Db 41 SerProProSerProProSerAsnSerThrThrThrProProProAlaSerAlaPro 60
QY 554 ATGCCTACTTCTCCCGTGGTCAACCCGCGCCGCCACGCAGCTGTCGGGACC----- 506
Db 61 ProProThrThrProSer-SerProProProProSerThrThrThrProProProPr 80
QY 505 -ACGGAATAACGACGCGCCTCGCGCGCGCCGCCCTCCCTCTCAGCCCGACCCGGA 447
Db 80 oSerSerArgSerThrProSerAlaProProProSerProProThrProSerThrProGl 100
QY 446 CCGGCAGCGTGCACGCATCGGTATGCCGCGGACGCGCGACCCCGCC---CCCTCTCCCTT 390
Db 100 ySerPro-----ProProLeuProGlnProSerProProAlaProTh 114
QY 389 ACTCCGGGGACTTCGGCTCCCGGCCACACCCCTCCCTCGCCACCGCTCCGTC-- 332
Db 114 xThrProGlySerProProAlaProValThrProProThrArgAsnProProProSerVa 134
QY 331 ----GGCCGATGGCGCGCCTTCTTCTCGCGAGATTAGACGATCGCGAGACCCGGA 276
Db 134 lProGly-----ProProSer----- 139
QY 275 AGTTATCTTTTTCACCTCTGTCCGAGTCCTGGGTAGTGGCGGTTCCTCCGGAAGAGGC 216
Db 140 -----AsnProSerArgGluGly-----GlySerProArgProProSe 152
QY 215 GGAGCCCGGAGTCTCAGAGCCCGCCCGCTCTGCGCTGCCCTCTGGG 170
Db 152 rSerProSerProProSerProSerSerAspGlyLeuSerThrGly 167

RESULT 12
US-10-142-515-11
; Sequence 11, Application US/10142515
; Publication No. US20030078399A1
```

```
; GENERAL INFORMATION:
; APPLICANT: SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH
; APPLICANT: Lloyd, Kenneth O.
; APPLICANT: Yin, Beatrice W.T.
; TITLE OF INVENTION: Nucleic Acid Sequence Encoding Ovarian Antigen, CA125, and Use:
; FILE REFERENCE: 649-A-US
; CURRENT APPLICATION NUMBER: US/10/142,515
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/290,480
; PRIOR FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 5877
; TYPE: PRT
; ORGANISM: Human Being
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(5877)
; OTHER INFORMATION: Amino acid sequence of MUC16B
US-10-142-515-11

Alignment Scores:
Pred. No.:      0.0404      Length:      5877
Score:          141.50      Matches:      148
Percent Similarity: 31.86%      Conservative: 76
Best Local Similarity: 21.05%      Mismatches: 266
Query Match:    2.57%      Indels:      213
DB:              15      Gaps:      27

US-10-001-857-42 (1-3096) x US-10-142-515-11 (1-5877)
QY 2237 CTTCCATTATCCTTTCTCTGCCATTTGAGAGCCATCGGCACGACTCAATGTTGACATCA 2178
Db 713 LeuThrLeuAspThrSerThrThrPheMetSerGlyThrHis-----Ser 727
QY 2177 ACCATGCGTAAAGGAATTCAGAGAGATACCAATATATGTAATAGTACTCGTGCTACTGT 2118
Db 728 ThrAlaSerGlnGlyPheSerHisSerGlnMetThrAlaLeuMetSerArgThrProGly 747
QY 2117 AGAGTTCCAATTCAAAGCCACTTAGAAGGTACTGTATCATATAATGCGAAGGTTATGGTAAA 2058
Db 748 GluValProTrpLeuSerHisProSer-----ValGluGluAlaSerSer 762
QY 2057 GGACCCAGGTACCTAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTCGAGCCCTGTTATG 1998
Db 763 AlaSerPheSerLeuSerSerProValMetThrSerSerSerProValSerSerThrLeu 782
QY 1997 TGTGAAGCGTGCATCAACCTTCTCTGCCTATTAAATAATAAATGTATAAACTCATCCTG 1938
Db 783 ProAspSerIleHisSerSerSerLeuProVal----- 793
QY 1937 CAAGGTGGCAAATTCCTCAAGAATATGACCAAGCTTATCTCTGTCTGAGCCCTGTTATG 1878
Db 794 -----ThrSerLeuLeuThr---SerGlyLeuValLys 803
QY 1877 TCCATGGATCTGAATAAGACTACAGAAATGGCCGAACACAGTGAATAACAAAGGAGTCGAT 1818
Db 804 ThrThrGluLeuLeuGlyThrSerSerGluProGluThrSerSerProProAsnLeuSer 823
QY 1817 ACAGTCCTTAGCCTGGTGATTATTATATAGGTAGCACTTGGGGAAAGCACTCGGAGGAT 1758
Db 824 SerThrSerAlaGluIleLeuAlaThrThrGluValThrThrAspThrGluLysLeuGlu 843
QY 1757 CTGACAAAAGACCCGAGTGCACTTTTCACCATGCTTCGCATGAGATGAGTTCCTCAAGACC 1698
Db 844 MetThrAsnValValThrSerGlyTyrThrHisGluSer-----ProSerSer 859
QY 1697 TTTTGTATTACC-----ACCAGGAAAAGTGGTTTGTAAACAGAGATCTTGAAGAAGACACAT 1644
Db 860 ValLeuAlaAspSerValThrThrLysAlaThrSerSerMetGlyIleThrTyrProThr 879
QY 1643 GGTGACTGTTCACTAAATTCA-----CAGAAAAAATCCAGGATACAAATGT 1599
```





```
; APPLICANT: James E.Bear
; APPLICANT: Jurgen Wehland
; APPLICANT: Joseph Loureio
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; FILE REFERENCE: M0656/7064 (HCL)
; CURRENT APPLICATION NUMBER: US/09/823,240
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-823-240-2
```

```
Alignment Scores:
Pred. No.: 0.0267 Length: 802
Score: 140.00 Matches: 49
Percent Similarity: 36.24% Conservative: 5
Best Local Similarity: 32.89% Mismatches: 58
Query Match: 2.55% Indels: 37
DB: 9 Gaps: 6
```

US-10-001-857-42 (1-3096) x US-09-823-240-2 (1-802)

```
QY 535 TCACCCGCGCGCCACGCACTGTCCGGACCGAAATAACGACGCCGCTCGCCGCGG 476
      ||||| ||| ||| ||||| ||||| ||||| |||||
Db 532 SerProThrProGlnGlyLeuVal-----LeuGlyProProAlaProPro 546

QY 475 CCGCCCCCTCCCTCTCAGCCAGCCGACCGGAGCGTGACGCAATGCGTATGCCCGGG 416
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 547 ProProProProLeuProSerGlyProAlaTyrAlaSerAlaLeuProProProGly 566

QY 415 ACGCGGACCCCGCCCTCTCCCTTACTCCGGGACTTCGGCGTCCCGGCCACACCC 356
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 567 -----ProProProProProLeuProSerThrGly-----ProProPro 580

QY 355 CCTCCCTCGCCACCGCTCCGTTCCGGCGATGGCGGCCGCTTCTCTTCGCGAGATTA 296
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 581 ProProProProProProLeuProAsnGlnAlaProPro----- 595

QY 295 GAGACGATCGGAGACCGGAAGTTATCTTTTTCACCCCTCTGTGGAGTCTGGGTAGTG 236
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 596 -----ProProProProAlaProProLeuProAlaSerGlyIlePhe 610

QY 235 GCGGTTCCCGAAAGAG-----GCGGAGCCCGGA 206
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 611 SerGlySerThrSerGluAspAsnArgProLeuThrGlyLeuAlaAlaIleAlaGly 630

QY 205 GTCTCAGAGCCCGCCCTGTGCGTGCCTCTGGGAATTTCTTAGAAAGGTGACGCGAA 146
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 631 AlaLysLeuArgLysValSerArgValGluAspGlySerPhe-----Pro 645

QY 145 GCGGAAGGAAACCGTGAGGTTTG 121
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 646 GlyGly-GlyAsnThrGlySerVal 653
```

RESULT 15

```
US-10-196-935A-4
; Sequence 4, Application US/10196935A
; Publication No. US20030082720A1
; GENERAL INFORMATION:
; APPLICANT: Lifton, Richard P
; APPLICANT: Wilson, Frederick H
; APPLICANT: Choate, Keith
; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Nelson-Williams, Carole
; TITLE OF INVENTION: COMPOSITIONS METHODS AND KITS RELATING TO TREATING AND DIAGNOSING
; FILE REFERENCE: 044574-5113
```

```
; CURRENT APPLICATION NUMBER: US/10/196,935A
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: US 60/306,084
; PRIOR FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-196-935A-4
```

```
Alignment Scores:
Pred. No.: 0.0341 Length: 1243
Score: 139.50 Matches: 61
Percent Similarity: 35.24% Conservative: 13
Best Local Similarity: 29.05% Mismatches: 63
Query Match: 2.54% Indels: 73
DB: 15 Gaps: 10
```

US-10-001-857-42 (1-3096) x US-10-196-935A-4 (1-1243)

```
QY 535 TCACCCGCGCGCCACGCACTGTC-----CGGACACGGAATAACGACGCCGCTCGG 482
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4 SerProAlaThrGluThrThrValLeuMetSerGlnThrGluAlaAspLeuAlaLeuArg 23

QY 481 CCGCCGCGCGCC-----CCTCCCTCTCAGCCCCAGCCCCGACGGCAG 440
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24 ProProProProLeuGlyThrAlaGlyGlnProArgLeuGlyProProProArgAla 43

QY 439 CGT-----GCACGCATGCGTATGCCCGGACGCGC----- 410
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44 ArgArgPheSerGlyLysAlaGluProArgProArgSerSerArgLeuSerArgArgSer 63

QY 409 -----GACCCGCGCCCTCTCTCC 392
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 64 SerValAspLeuGlyLeuLeuSerSerTrpSerLeuProAlaSerProAlaProAspPro 83

QY 391 -----TTACTCCCGGGGACTTCGCGTCCCGGCCACACCCCTCTCTCC 350
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84 ProAspProProAspSerAlaGlyProGlyProAlaArgSerPro-----ProProSer 101

QY 349 TCGCCACCGCTCCGTTCCGCCGATGGCGCGCCTTCTTCTCGCGAGATTAGAGACG 290
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 102 SerLysGluProProGluGlyThrTrpThrGluGlyAlaProValLysAlaAlaGluAsp 121

QY 289 ATCGCGAGACCGGAAGTTATCTTTTTCACCCCTCTGTCGGAGTCTCTGGGTAGTGGCGGT 230
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 SerAlaArgProGlu-----LeuProAspSerAlaValGlyProGly 135

QY 229 TCCCGGAAAGCGGAGCCCGAGTCTCAGAGCCCGCCCGTCTGCGTGCCTCTCTGGG 170
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 136 SerArgGlu-----ProLeuArgValProGluAla 145

QY 169 AATTTCTTAGAAAGGTGACCGGAAGCGGAAGAAACCGTGGAGGTTTCCGTAGACCTTA 110
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 146 ValAlaLeuGluArgArgGlu----- 153

QY 109 GTGGTTGCGAGCACGAACGAGCAGCCCAAGCTCCGAGAGCAACAGCGGTGCTACTGCTGCCT 50
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154 -----GlnGluGluLysGluAspMetGluThrGlnAlaValAlaThrSerPro 169

QY 49 GACAACGGATTATG---TTTCTTATACAG 23
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 170 AspGlyArgTyrLeuLysPheAspIleGlu 179
```

Search completed: November 25, 2003, 03:26:56  
Job time : 155 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2003, 21:46:55 ; Search time 6196 Seconds  
(without alignments)  
12144.393 Million cell updates/sec

Title: US-10-001-857-42  
Perfect score: 3096  
Sequence: 1 ttcttcacgaaactcccagg.....atttcttgacaaaaaaaaa 3096

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	795	25.7	1033	11 BC027201	BC027201 Mus muscu
2	777.8	25.1	973	11 AK021042	AK021042 Mus muscu
3	705	22.8	722	12 BM979022	BM979022 UI-CF-DU1
4	696.8	22.5	854	10 BF969365	BF969365 602271630

C	5	658.6	21.3	814	14	CB169246	CB169246 VBB603020
	6	653.8	21.1	3071	11	AK050805	AK050805 Mus muscu
	7	653.6	21.1	768	10	BG547889	BG547889 602576187
	8	648	20.9	957	13	BQ965085	BQ965085 AGENCOURT
C	9	641.4	20.7	828	9	AI790514	AI790514 ul02b03.x
C	10	610	19.7	616	12	BM979511	BM979511 UI-CF-DU1
	11	599.2	19.4	900	14	CA977646	CA977646 AGENCOURT
	12	598.8	19.3	697	12	BI602827	BI602827 603247187
C	13	597	19.3	598	9	AW007583	AW007583 wt02f11.x
	14	593.8	19.2	718	13	BU058931	BU058931 UI-M-FR0-
	15	586	18.9	972	10	BG112812	BG112812 602285005
	16	584.8	18.9	2958	11	BC043701	BC043701 Mus muscu
	17	575.2	18.6	763	13	BU613022	BU613022 UI-M-FR0-
	18	572.4	18.5	952	13	BQ925670	BQ925670 AGENCOURT
C	19	560	18.1	560	9	AA888870	AA888870 ak23f02.s
	20	555.6	17.9	927	11	BC010475	BC010475 Mus muscu
	21	551.2	17.8	1117	13	BU525571	BU525571 AGENCOURT
	22	548.4	17.7	873	9	AU081000	AU081000 AU081000
	23	542	17.5	735	13	BU057675	BU057675 UI-M-FR0-
	24	538.6	17.4	654	10	BF468052	BF468052 UI-M-CD0-
	25	537	17.3	537	12	BM699453	BM699453 UI-E-DX1-
	26	526.8	17.0	666	13	BQ443173	BQ443173 UI-M-EV0-
	27	523.2	16.9	637	14	CD372119	CD372119 UI-R-GO0-
	28	507.4	16.4	509	10	BG612374	BG612374 602640115
	29	506.8	16.4	906	12	BI658829	BI658829 603301980
	30	503.2	16.3	1114	10	BG245408	BG245408 602354951
	31	499.6	16.1	640	13	BQ443252	BQ443252 UI-M-EV0-
	32	496.6	16.0	1049	13	BQ885743	BQ885743 AGENCOURT
	33	495.6	16.0	631	14	CB418817	CB418817 591640 MA
	34	491.4	15.9	690	10	AW975643	AW975643 EST387752
C	35	490.8	15.9	901	12	BI731865	BI731865 603353606
	36	487	15.7	494	12	BM510376	BM510376 ij40g03.y
	37	482.2	15.6	721	9	AJ453000	AJ453000 AJ453000
	38	480.2	15.5	615	14	CA749517	CA749517 UI-M-FY0-
C	39	472.6	15.3	479	9	AI339059	AI339059 qq29e09.x
C	40	470.4	15.2	485	9	AW090031	AW090031 xd01a08.x
C	41	465	15.0	480	14	CB242727	CB242727 UI-CF-FN0
	42	464.8	15.0	556	10	BG062635	BG062635 L0955F03-
	43	463.2	15.0	558	14	CA555718	CA555718 K0201G09-
	44	461.6	14.9	569	12	BI682349	BI682349 463408 MA
	45	461.4	14.9	784	13	BU260643	BU260643 603503268

ALIGNMENTS

RESULT 1  
BC027201  
LOCUS BC027201 1033 bp mRNA linear HTC 20-SEP-2002  
DEFINITION Mus musculus, Similar to corneal wound healing related protein, clone IMAGE:4483189, mRNA.  
ACCESSION BC027201  
VERSION BC027201.1 GI:20070883  
KEYWORDS HTC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Strausberg,R.  
AUTHORS Direct Submission  
TITLE Submitted (04-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov  
COMMENT Contact: MGC help desk  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center



Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaïdo, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weltz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.

**TITLE** Functional annotation of a full-length mouse cDNA collection  
**JOURNAL** Nature 409 (6821), 685-690 (2001)

MEDLINE	21085660
PUBMED	11217851
REFERENCE	5

**THE FANTOM CONSORTIUM AND THE RIKEN GENOME EXPLORATION RESEARCH GROUP PHASE I & II TEAM.**

**TITLE**  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL  
NATURE 420, 563-573 (2002)  
REFERENCE  
6 (bases 1 to 973)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,  
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,  
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,  
Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,  
Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,  
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,  
Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,  
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,  
Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,  
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,  
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222

COMMENT  
Fax:81-45-503-9216)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through two rounds of normalization to Rot = 20.0 and subtraction to Rot = 478.8. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGATCTCGACTTAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

**FEATURES**

```
1. 973
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:C030004C14"
/db_xref="MGI:1912689"
/db_xref="taxon:10090"
/clone="C030004C14"
```

FEATURES  
 sou

```

/misc_feature
24. _>971
/note="CORNEAL WOUND HEALING RELATED PROTEIN homolog
[Rattus norvegicus] (SPTR|Q9JI01, evidence: FASTA,
96.2%ID, 43.5%length, match=948)
putative"
/db_xref="MGI:1925939"
BASE COUNT      285 a      202 c      250 g      236 t
ORIGIN

```

Query Match 25.1%; Score 777.8; DB 11; Length 973;  
Best Local Similarity 87.9%; Pred. No. 5.3e-130;  
Matches 848; Conservative 0; Mismatches 117; Indels 0;

541	QY	GGAGAAGTAGGCATAAATGGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGGATGGGAG	600
9	Db	GGCGGCTAGGCATCATGGTTATGAAAGCTACTGTAGACGACGATGCTTCGGGATGGGAG	68
601	QY	CTCAGTATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCCAAAGAT	660
69	Db	CTCGGGTCCCGGAAAAAATGGAAAAAAGTAGCACAAAGCTGGGTGGACATAACCCAGGAT	128
661	QY	TTTGAAGAAGCTTGTCGAGAATTAAAGTTGGGAGAACTACTTTCATGATAAGCTATTTTGGT	720
129	Db	TTTGAAGATGCTTGTCGAGAGCTGAAGTTGGGAGAACTGCTTCATGATAAGCTGTTTGGT	188
721	QY	CTTTTGAAGCCATGTCCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATT	780
189	Db	CTTTTGAAGCCATGTCCTGCTATTGAAATGATGGATCCTAAGATGGATGCCGTATGATC	248
781	QY	GGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT	840
249	Db	GGGAACCAAGTGAATAGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACCAT	308
841	QY	AAAAATTAAAGATCTCACCTTGCTGAACTGATAGGGATTATGGATACATGTTTTCGTGT	900
309	Db	AAAAATTAAAGACCTCAGCCTGCCTGAACTGATAGGAATAATGGACACCTGTTCTGCTGT	368
901	QY	TTGATAACGTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTTACGTGCTTTTACATT	960
369	Db	TTGATCACATGGCTCGAAGGCCATTCTTGGCACACACAGGGTTTACGTGCTTTACATT	428
961	QY	CATAATCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGGCTCTGGGAACTTTGAAA	1020
429	Db	CATAATCCGACTTCATAGAAGATCCTGCCATGAAAGCTTTTGTTCGGGAACTCTTGAAG	488
1021	QY	ATCTGTGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTT	1080
489	Db	ATTGGTACATTGCACGGAAAAAGTAAATAAAGCTGCTGTTTTTCGAAGAGGAAGATTTT	548
1081	QY	CAGTCAATGACTTATGGATTTAAAAATGGCTAACAGTGTGACAGATCTTCAGATTACAGGC	1140
549	Db	CAGTCAATGACATACGGATTTAAAAATGGCCAACAGTGTGACAGATCTCCGAGTTACAGGG	608
1141	QY	ATGCTAAAAGATGTGGAGGATGACATGCAAAAGAGAGTAAAGAGTACTCGAAAGTCGACAA	1200
609	Db	ATGATAAAGGATGTGGAAGATGATCTGCAAAAGCGGAGTAAAGAGTACTTGAAGTCGACAA	668
1201	QY	GGAGAAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGTTTAGCAGTATTCAGC	1260
669	Db	GGAGAAGAGAGAGATCCAGAGGTGAACTAGAACACCCAGCAGTGTGTTGGCAGCATTCAGC	728
1261	QY	AGAGTGAAATTTTACTCGTGTGTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAAGAGACC	1320
729	Db	AGAGTGAAAGTTACCGCAGTGTGCTCACAGTGTCTATAGCCTTTTACTAAGAAAAGAGACC	788
1321	QY	AGTGCTGTTGCAGAAGCTCAAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATT	1380
789	Db	AGTGCTGTTGCAGAGGCTCAGAAAATGATGGTCCAGGCGGACACCTTCTTTCTGCCATT	848

```
QY 1381 CATAATTTCATCATGGCATCCAGGCCAGGATGATGATGATACTACAAAAGGAGATCATCCCA 1440
Db 849 CACACCTCATTCACACACGGCATCCAGGCTCAGAAATGGCACTACCAAAGGAGACCATCCA 908

QY 1441 ATTATGATGGGTTTTGAACCCCTTGTGAACACGAGGCTACTTCCACCTACCTTCCCTCGA 1500
Db 909 ATTATGATGGGTTTTGAGCCCTTGTTAAACGAGAGATTACTTCCACCCACCTTCCCTCGC 968

QY 1501 TATGC 1505
Db 969 TATGC 973

RESULT 3
BM979022/c
LOCUS
DEFINITION
  UI-CF-DU1-adl-c-13-0-UI.s1 UI-CF-DU1 Homo sapiens cDNA clone
  UI-CF-DU1-adl-c-13-0-UI 3', mRNA sequence.
ACCESSION
  BM979022
VERSION
  BM979022.1 GI:19599047
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 722)
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
JOURNAL
  Genome Res. 6 (9), 791-806 (1996)
MEDLINE
  9704477
PUBMED
  8889548
COMMENT
  Contact: McCray, PB
  McCray Lab
  University of Iowa
  2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
  Tel: 319 356 4866
  Fax: 319 356 7171
  Email: paul-mccray@uiowa.edu
  Tissue Procurement: Dr. M. J. Welsh, University of Iowa
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Researchers may obtain clones from Research
  Genetics (www.resgen.com) or from Open Biosystems
  (www.openbiosystems.com).
  Seq primer: M13 FORWARD
  POLYA=Yes.
  Location/Qualifiers
    1..722
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="UI-CF-DU1-adl-c-13-0-UI"
    /tissue_type="Primary Lung Epithelial Cells"
    /dev_stage="Adult"
    /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
    /clone_lib="UI-CF-DU1"
    /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
    modified polylinker; Site 1: EcoR I; Site 2: Not I;
    UI-CF-DU1 is a normalized_cDNA library containing the
    following tissue(s): Primary Lung Epithelial Cells The
    library was constructed according to Bonaldo, Lennon and
    Soares, Genome Research, 6:791-806, 1996. First strand
    cDNA synthesis was primed with an oligo-dT primer
    containing a Not I site. Double stranded cDNA was ligated
    to an EcoR I adaptor, digested with Not I, and cloned
    directionally into pT7T3-pac vector. The oligonucleotide
    used to prime the synthesis of first-strand cDNA contains
    a library tag sequence that is located between the Not I
    site and the (dT)18 tail. The sequence tag for this
    library is GGCTGTAGGC.
    TAG_LIB=UI-CF-DU1
```

```

TAG TISSUE=Lung Epithelial Cells Tissue nos 359-368
TAG_SEQ=GGCTGTAGGC"
BASE COUNT 215 a 147 c 124 g 236 t
ORIGIN
Query Match 22.8%; Score 705; DB 12; Length 722;
Best Local Similarity 99.9%; Pred. No. 7.2e-117;
Matches 716; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 2380 CAAAGTACGTAAACCGAAGTTTGAGCTTGATAGTGAAACAAGTTCGGTATGAACACACAGGTT 2439
Db 722 CAAAGTACGTAAACCGAAGTTTGAGCTTGATAGTGAAACAAGTTCGGTATGAACACACAGG-T 664

QY 2440 TGCTCCATTCAACAGTGTGATGACCCCGCCGCGCAGTGCACTACTTACAGTTCAAGGAAAT 2499
Db 663 TGCTCCATTCAACAGTGTGATGACCCCGCCGCGCAGTGCACTACTTACAGTTCAAGGAAAT 604

QY 2500 GTCTGACCTCAATAAATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAG 2559
Db 603 GTCTGACCTCAATAAATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAG 544

QY 2560 TAAGCACTTTCAACAGGCAAAAATGATATTGGAAAATATTCCTAACCCGACCATGAGGT 2619
Db 543 TAAGCACTTTCAACAGGCAAAAATGATATTGGAAAATATTCCTAACCCGACCATGAGGT 484

QY 2620 TAATAGAAATTTTAAAGGTTGCCAAACCCCAACTTTTGTTGTTATGAAGTTATTGGCAGGAGG 2679
Db 483 TAATAGAAATTTTAAAGGTTGCCAAACCCCAACTTTTGTTGTTATGAAGTTATTGGCAGGAGG 424

QY 2680 ACACAAAAGGAATCTAAAGTTCCTCCTGAATTTGATTTCTCTGCTCATAAATATTTTCC 2739
Db 423 ACACAAAAGGAATCTAAAGTTCCTCCTGAATTTGATTTCTCTGCTCATAAATATTTTCC 364

QY 2740 TGTTGTGAAAACCTTGTTTGAGAGAGACTGGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTT 2799
Db 363 TGTTGTGAAAACCTTGTTTGAGAGAGACTGGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTT 304

QY 2800 TCAGACCCCAACTCTTAGAGGSCACATCACCAAGGCTCCACATCACGGAAGTGAGATGGAT 2859
Db 303 TCAGACCCCAACTCTTAGAGGSCACATCACCAAGGCTCCACATCACGGAAGTGAGATGGAT 244

QY 2860 TTCTTTGGGTAACAACACTCAITATAAGGAATACCTTTTAGTTTGACAGCCTTATATGACATGA 2919
Db 243 TTCTTTGGGTAACAACACTCAITATAAGGAATACCTTTTAGTTTGACAGCCTTATATGACATGA 184

QY 2920 ATGAAAACCTGCTGTTTTTAAAGTGGTTTATTATGTTCCATGGAAGAAACTGGTCTTATTGA 2979
Db 183 ATGAAAACCTGCTGTTTTTAAAGTGGTTTATTATGTTCCATGGAAGAAACTGGTCTTATTGA 124

QY 2980 ATGCATTGATGAACGTTTATATGGTTTTTATTACAGATTTAAATCACAAATCATTTTTTATGA 3039
Db 123 ATGCATTGATGAACGTTTATATGGTTTTTATTACAGATTTAAATCACAAATCATTTTTTATGA 64

QY 3040 ATGATTGAGTGAAAATAGTGTTTATAAAGGTTAATAAATTTCTTGACAAAAAATAA 3096
Db 63 ATGATTGAGTGAAAATAGTGTTTATAAAGGTTAATAAATTTCTTGACAAAAAATAA 7

RESULT 4
BF969365
LOCUS
DEFINITION
  BF969365
  602271630F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4359702 5',
  mRNA sequence.
ACCESSION
  BF969365
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 854)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
```

JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC

cdNA Library Preparation: Life Technologies, Inc.  
cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM999 row: 1 column: 07  
High quality sequence stop: 785.

FEATURES  
source Location/Qualifiers  
1..854  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4359702"  
/tissue\_type="adrenal cortex carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC\_84"  
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT 273 a 158 c 186 g 237 t  
ORIGIN

Query Match 22.5%; Score 696.8; DB 10; Length 854;  
Best Local Similarity 94.5%; Pred. No. 2.1e-115;  
Matches 809; Conservative 0; Mismatches 37; Indels 10; Gaps 8;

QY 2245 GAAAGGCCGTAGTAGTAAAAAACAAG-AAAAAAGAAAGTTCCGCCATTGAGCCGAG 2303  
Db 1 GAAAGGCCGTAGTAGTAAAAAACAAGAAAAAAGAAAGTTCCGCCATTGAGCCGAG 60

QY 2304 AGATCACAATGAGCCAAAGCATATCAGAACATGTGTGCTGGAATGTTTAAACCATGGTAG 2363  
Db 61 AGATCACAATGAGCCAAAGCATATCAGAACATGTGTGCTGGAATGTTTAAACCATGGTAG 120

QY 2364 CATTTGACATGAGCGCAAGTACGTAAACCGAAGTTTGAGCTTGATAGTGAACAAAGTTC 2423  
Db 121 CATTTGACATGAGCGCAAGTACGTAAACCGAAGTTTGAGCTTGATAGTGAACAAAGTTC 180

QY 2424 GGTATGAACACAGGTTTGCTCCATTCAACACTGTGATGACCCCGCCGCGCAGTGCACACT 2483  
Db 181 GGTATGAACACAGGTTTGCTCCATTCAACACTGTGATGACCCCGCCGCGCAGTGCACACT 240

QY 2484 TACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTCCTCAGTCTCCTGAAC 2543  
Db 241 TACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTCCTCAGTCTCCTGAAC 300

QY 2544 TGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCAAAAATGATATTGGAAAAATATTCCTA 2603  
Db 301 TGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCAAAAATGATATTGGAAAAATATTCCTA 360

QY 2604 ACCCGGACCATGAGGTTAATAGAATTTTAAAGTTGCCAAACCCCACTTTGTGGTTATGA 2663  
Db 361 ACCCGGACCATGAGGTTAATAGAATTTTAAAGTTGCCAAACCCCACTTTGTGGTTATGA 420

QY 2664 AGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCCTCCTGAATTTGATTTCCTCG 2723  
Db 421 AGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCCTCCTGAATTTGATTTCCTCG 480

QY 2724 CTCATAAATATTTTCCTGTTGTGAACTTGTGTGAGAGAGACTGGGGAGGTGGCCATAAA 2783  
Db 481 CTCATAAATATTTTCCTGTTGTGAACTTGTGTGAGAGAGACTGGGGAGGTGGCCATAAA 540

QY 2784 GGGGCAGAGTCTTCTTTTCAGACCCCACTTTAGAGGGCACATACACAGGCTCCACATCAC 2843  
Db 541 GGGGCAGAGTCTTCTTTTCAGACCCCACTTTAGAGGGCACATACACAGGCTCCACATCAC 600

QY 2844 GGGAAAGTGAGATGGATTCTTGGGTAACAACCTCATTATAAGGAATACTTTAG-TTTGAC 2902  
Db 601 GGGAAAGTGAGATGGATTCTTGGGTAACAACCTCATTATAAGGAATACTTTAGTTTGAC 660

QY 2903 AGCCTTATATGACA-TGAATGAAAACTGCTGTTTTAA-AGTGGTTTATTATGTTCCAT-G 2959  
Db 661 AGCCTTATATGACATTGAATGAAAACTGCTGTTTTAAACAGTGGTCTTTATGTTCCCTGG 720

QY 2960 GAAGAAACTGGTCTTATTGAATGCATTGATGACGTTATATGTTTATTACAGATTAA 3019  
Db 721 GACGAAACTGGTCTTATTGAATGGCTTGATGAACG-TATATGGGTTTATTACCGATTAA- 778

QY 3020 TCACAAATCATTTTATTATGAATGATTGAGTGAATAAGTAGTGTTTATAAAGTTATAAATT 3079  
Db 779 -TCCCAATCATTTTATTATGAATGATTG-GTGCACATCGGTTTAAACGTGATAAATTCTTG 835

QY 3080 TCTTGACAAAAA 3095  
Db 836 CGCAAAAAACAAAAA 851

RESULT 5  
CB169246/c  
LOCUS  
DEFINITION VBB603020215.R1 CSEQFXN41 testes Bos taurus linear EST 30-JAN-2003  
ACCESSION CB169246  
VERSION CB169246.1 GI:28155372  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 814)  
AUTHORS Adelson,D.L. and Gill,C.A.  
TITLE Bovine ESTs (Adelson and Gill)  
JOURNAL Unpublished  
COMMENT Contact: David L. Adelson  
Animal Breeding and Genetics  
Texas A&M University  
Animal Science Dept., TAMU-2471, College Station, TX 77843-2471,  
USA  
Tel: 9798452616  
Fax: 9798456970  
Email: david.adelson@tamu.edu.  
Location/Qualifiers  
1..814  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/tissue\_type="testes"  
/clone\_lib="CSEQFXN41 testes"  
/note="Organ: testes; Vector: pBluescript SK+; Site\_1: NotI; Site\_2: EcoRI; sequence 5' of the insert (5'-NNN...NNNinsert)  
GCGAATTGGAGCTCCACCGGCGGCGGCGGCTCGAG. Sequence 3' of the inserts (AAGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAG. normalized Rd 1 library, sequenced 3' with M13R primer."

BASE COUNT 254 a 161 c 173 g 226 t  
ORIGIN

Query Match 21.3%; Score 658.6; DB 14; Length 814;  
Best Local Similarity 90.3%; Pred. No. 1.6e-108;  
Matches 758; Conservative 0; Mismatches 54; Indels 27; Gaps 4;

QY 1348 ATGTTTCAAGCAGCAGATCTTCTTTCTGCCATTTCATTAATTCATGTCATGCCATCCAG 1407  
Db 814 ATGTTTCAAGCAGCAGATCTTCTTTCTGCCATTTCATTAATTCATGTCATGCCATCCAG 755

QY 1408 GCCCAGATGATACT-ACAAAAGGAGATCATCCAATTATGATGGGTTTGAACCCCTTGT 1466  
Db 754 GCCCAGATGACACTAACAAAAGGAGACCATCCCATTATGATGGGTTTGAAGCCACTTGT 695

Qy	1467	GAACGAGGGCTACTTCCACCTACCTTCCCTCGATATGCAAAAATAAATAAAAGGAAGA	1526
Db	694	TAACCAAGACTACTTCCACCCACCTTCCCTCGATATGCAAAAATAAATAAAAGAGAAGA	635
Qy	1527	AATGGTGAATATTTTGCAAGATTATPAGATAGAATAAAAACTGCTGTGAGGTTGTGAA	1586
Db	634	AATGGTCAACTATTTTGAAGATTGATAGATAGAATAAAAACTGCTGTGAGGTGGTCAA	575
Qy	1587	TTTAACAAATTTACATTGTATCCTGGATTTTTTCTGTGAATTTAGTGAACAGTCACCATG	1646
Db	574	TTTAACAAATTTACATTGTATCTTGGATTTTTTCTGTGAATTTAGTGAACAATCACCTTG	515
Qy	1647	TGTTCTTTCAAGATCTCTGTTACAAACCACTTTCCTGGTGGATAACAAAAAGGTCCTTGG	1706
Db	514	TGTTCTTTTCGAGATCTCTATTACAAACCACCTTTCCTGGTGGATAACAAAAAGGTCCTTGG	455
Qy	1707	AACTCATCTCATGCAAGACATGGTGAAGATGCACCTTCGGTCTTTTGTTCAGATCCTCCGA	1766
Db	454	AATCCATCTCATGCAAGACATGGTGAAGATGCTCTTCGGTCTTTGTTCAGTCCTCCGG--	396
Qy	1767	GTGCTTTCCCCCAAGTGCTACCTATATATAATCACCAGGCTAAGGACTGTATCGACTCC	1826
Db	395	-TGCTCTCCCCAAAGTGCTGCCTATATATAATCACCAGGCTAAGGACTGTATTGACTCT	337
Qy	1827	TTTGTTACTCACTGTGTTCCGGCCATTCCTAGTCTTATTTCAGATCCATGGACATAACAGG	1886
Db	336	TTTGTTACTCACTGTGTTCCGGCCATTCCTAGTCTTATTTCAGATCCATGGACATAACAGG	277
Qy	1887	GCTCGACAGAGAGATAAGCTTGGTCAATTCCTTGAGGAATTTTGCCACCTTGCAGGATGAG	1946
Db	276	GCTCGGCAGAGAGATAAACTCGGTCAATTCCTTGAGGAATTTGTACCTTGCAGGATG--	219
Qy	1947	TTTATGACATTTTATTTTAATAGGCAGAGAAGGTTGATGCAGCGCTTCACACCATGCTGT	2006
Db	218	-----AGGCAGAGAAGGTTGATGCAGCCCTTCATACTATGCTGT	180
Qy	2007	TGAAACAGGAACCCCAAGGCAACATTTTGGCTGTTTAGGTACCTGGTCCCTTTACCATA	2066
Db	179	TGAAACAGGAGCCCCACAGGCAGCATTTTGGCCTGTTTAGGTACCTGGTCCCTTTACCATA	120
Qy	2067	ACCTTCGCATTATGATACAGTACCTTCFAAGTGGCTTTTGAATTGGAACCTCTA-CAGTATG	2125
Db	119	ACCTTCGAATTATGATACAGTATCTCCTAAGTGGCTTCGAATTGGAGCTCTACCAGCATG	60
Qy	2126	CACGAGTACTATTACATATATTGGTATCTCTCTGAATTCCTTTACGCATGGTTGATGTC	2184
Db	59	CACGAGTATTATTACATATATTGGTATCTCTCTGAATTTCTTTATGCTGGTTGATGTC	1
RESULT 6			
AK050805			
LOCUS		3071 bp mRNA linear HTC 05-DEC-2002	
DEFINITION		Mus musculus 9 days embryo whole body cDNA, RIKEN full-length enriched library, clone:D030020M24 product:CORNEAL WOUND HEALING RELATED PROTEIN homolog [Rattus norvegicus], full insert sequence.	
ACCESSION		AK050805	
VERSION		AK050805.1 GI:26094130	
KEYWORDS		HTC; CAP trapper.	
SOURCE		Mus musculus (house mouse)	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS		1	
TITLE		High-efficiency full-length cDNA cloning	
JOURNAL		Meth. Enzymol. 303, 19-44 (1999)	
MEDLINE		99279253	
PUBMED		10349636	
REFERENCE		2	
AUTHORS		Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	
TITLE		Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159

3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

TITLE

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

4 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851

TITLE

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3071)  
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
Direct Submission

TITLE

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.

TITLE

JOURNAL  
COMMENT

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.



||||| 181 GAACATGTGCTGGAAATGTTAAAAACCATGGTAGCATTTGACATGGACGGCAAAAGTACG 240

||||| QY TAAACCGAAGTTTGAGCTTGATAGTGAACAAGTTTCGGTATGAACACAGGTTTGTCTCCATT 2448

||||| Db TAAACCGAAGTTTGAGCTTGATAGCGAAACAAGTTTCGGTATGAACACACAGGTTTGTCTCCATT 300

||||| QY CAACAGTGTGATGACCCCGCCAGTGCACCTACTTACAGTTCGAAGAAATGTCTGACCT 2508

||||| Db CAACAGTGTGATGACCCCGCCAGTGCACCTACTTACAGTTCGAAGAAATGTCTGACCT 360

||||| QY CAATAA-ATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACT 2567

||||| Db CAATAACATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACT 420

||||| QY TTCAACAGGCAAAAATGATATTGGAATAATATTCCTAACCCGACCATTGAGTTTCATAGAA 2627

||||| Db TTCAACAGGCAAAAATGATATTGGAATAATATTCCTAACCCGACCATTGAGTTTCATAGAA 480

||||| QY TTTTAAAGGTTGCCAAACCCCACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAA 2687

||||| Db TTTTAAAGGTTGCCAAGACCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAA 540

||||| QY AGGAATCTAAAGTTCTCCTGAAATTGATTCTCTGCTCATAAATATTTTCCTGTTGTGA 2747

||||| Db AGGAATCTAAAGTTCTCCTGAAATTGATTCTCTGCTCATAAATATTTTCCTGTTGTGA 600

||||| QY AACTTGTTTGAGAGAGACTGGGGAGGTGGCCAT-AAAGGGGAGAGTCTTCTTTCAGACC 2806

||||| Db AACTTGTTTGAGAGAGACTGGGGAGGTGGCCATCAAACGGGGAGAGTCTTCTATCAGACC 660

||||| QY CAACTCTTAGAGGGCACATCACAGGCTCCACATCACGGGAAGTGAGATGGAATTCCTGG 2866

||||| Db CAAATCTTAGAGGGCCATTTCACGAGGCTCCA-ATCACGGGAAGTGAGATGGAATTCCTGGG 719

RESULT 8 BQ965085 957 bp mRNA linear EST 21-AUG-2002

LOCUS AGENCOURT 10052197 NIH\_MGC\_134 Mus musculus cDNA clone

DEFINITION IMAGE:6509503 5', mRNA sequence.

ACCESSION BQ965085

VERSION BQ965085.1 GI:22380563

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 957)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Dr. David Rowe  
cDNA Library Preparation: Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM14076 row: 0 column: 08  
High quality sequence stop: 632.  
Location/Qualifiers  
1. .957  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6509503"  
/tissue\_type="undifferentiated limb"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_134"  
/note="Vector: pCMV-SPORT6.1.ccdB; Site\_1: EcoRV; Site\_2:

NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH MGC Library."

BASE COUNT 274 a 198 c 253 g 232 t

ORIGIN

Query Match 20.9%; Score 648; DB 13; Length 957;  
Best Local Similarity 87.7%; Pred. No. 1.3e-106;  
Matches 719; Conservative 0; Mismatches 100; Indels 1; Gaps 1;

QY 541 GGAGAAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGGATGGGAG 600

Db 26 GCGCGGACGGCATCATGTTATGAAAGCTACTGTAGACGACGATGCTTCGGGATGGGAG 85

QY 601 CTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAAACTGGGTGGACATTACCCCAAGAT 660

Db 86 CTCGGGTCCCGGAAAAAATGGAAAAAGTAGCACAAAGCTGGGTGGACATAACCCAGGAT 145

QY 661 TTTGAAGAAGCTTGTGAGAAATTAAAGTTGGGAGAACTACTTTCATGATAAGCTATTTGGT 720

Db 146 TTTGAAGATGCTTGTGAGAGACTGAAAGTTGGGAGAACTGCTTCATGATAAGCTGTTTGGT 205

QY 721 CTTTGTGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATT 780

Db 206 CTTTGTGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATC 265

QY 781 GGAAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT 840

Db 266 GGGAAACCAAGTGAATAGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT 325

QY 841 AAAATTAAAGATCTCACCTTGCTGAACTGATAGGGATTATGGATACATGTTTGTCTGT 900

Db 326 AAAATTAAAGACCTCAGCCTGCTGAACTGATAGGAATATGACACACCTGTTTCTGCTGT 385

QY 901 TTGATAACGTGGTTAGAAAGGCCATTCACTGGCACAGACAGTATTTACGTGCCTTTACATT 960

Db 386 TTGATCACATGGCTCGAAAGGCCATTCTTTGGCACAGACAGTGTTCGTGCCTTTACATT 445

QY 961 CATAAATCCAGACTTTATAGAAAGATCCTGCTATGAAGGCTTTTGTCTCTGGGAATCTTTGAAA 1020

Db 446 CATAAATCCGACTTTCATAGAAGATCCTGCCATGAAAGCTTTTGTCTCTGGGAATCTTTGAAG 505

QY 1021 ATCTGTGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTTTGTGAAGAGGAAGATTTT 1080

Db 506 ATCTGCGACATTGCACGGGAAAAAGTAAATAAAGCTGCTGTTTGTGAAGAGGAAGATTTT 565

QY 1081 CAGTCAATGACTTATGGATTAAATGGCTAAACAGTGTACAGATCTTCGAGTTACAGGC 1140

Db 566 CAGTCAATGACATACGGATTAAATGGCTAAACAGTGTACAGATCTTCGAGTTACAGGC 625

QY 1141 ATGCTAAAAGATGTGGAGGATGACATGCAAGAGAGAGTAAGAGTACTCGAAGTCGACAA 1200

Db 626 ATGCTAAAAGGATGTGGAAGATGATCTGCAAAAGGCGAGTAAAGAGTACTCGAAGTCGACAA 685

QY 1201 GGAGAAAGAAAGAGATCCAGAAAGTTGAACCTAGAACACCAACAATGTTTAGCAGTATTCAGC 1260

Db 686 GGAGAAAGAGAGATCCCGAGGTGAACTAGAACACCAAGCAGTGTTCGGCAGCATTCAGC 745

QY 1261 AGAGTGAATTTTACTCGTGTGTACTGACAGTGTCTTATAGCCTTACTAAGAAAGAGACC 1320

Db 746 AGAGTGAAGTTTACGCGAGTGTGCTCACAGTGTCTATAGGCTTTACTAAG-AAGAGAAC 804

QY 1321 AGTGTGTGTGCAGAAAGCTCAAAAATTGATGGTTTCAAGCAG 1360

Db 805 AGTGTGTGTGCAGAAAGCTCAAAAATTGATGGTGGGCCAGCGG 844

RESULT 9 AI790514/c

LOCUS AI790514 828 bp mRNA linear EST 02-JUL-1999

DEFINITION ul02b03.x1 Sugano mouse kidney mkia Mus musculus cDNA clone

IMAGE:2064845 3', similar to WP:T23B12.4 CE14032 ;, mRNA sequence.

ACCESSION AI790514

VERSION AI790514.1 GI:5338230  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 828)  
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
TITLE The WashU-NCI Mouse EST Project 1999  
JOURNAL Unpublished  
COMMENT Other\_ESTs: ul02b03.y1  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:994033  
Possible reversed clone: similarity on wrong strand  
Seq primer: custom primer used  
High quality sequence stop: 509.  
FEATURES  
source  
1. .828  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2064845"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="Sugano mouse kidney mkia"  
/note="Organ: kidney; Vector: pME18S-FL3; Site 1: DraIII (CACTGTGTG); Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."  
BASE COUNT 208 a 197 c 177 g 240 t 6 others  
ORIGIN  
Query Match 20.7%; Score 641.4; DB 9; Length 828;  
Best Local Similarity 86.4%; Pred. No. 2e-105;  
Matches 716; Conservative 0; Mismatches 112; Indels 1; Gaps 1;  
QY 673 TGTGAGAAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTTGGTCTTTTGAAGCC 732  
Db 828 TGTGAGAGCTGAAGTTGGAGACTGCTTCATTGATAAGCTGTTGTTCTTTTGAAGCC 769  
QY 733 ATGTCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAAACCAAGTT 792  
Db 768 ATGTCTGCTATTGAAATGATGATTCNTAAGATGGATGCGGTATGATCNGAACCAAGTG 709  
QY 793 AATCGAAAAGTTCTCAATTTGAACAAGCTATCAAGGATGGCACTATTATAAATTAAAGAT 852  
Db 708 AATAGNAAAAGTTCTCAATTTGANCAAGCTATCAAGGATGGCAC-ATTAAATTTAAAGAC 650  
QY 853 CTCACCTTGCCTGAACTGATAGGATTAATGATACATGTTTGTGCTGTTGATAACGTGG 912  
Db 649 CTCAGCCTGCNTGAACTGATAGGAATAATGGACACCTGTTCTGCTGTTTGATCACAATGG 590

QY 913 TTAGAAGGCCATTCACTGSCACAGACAGTATTTTACGTGCTTTTACATTTCATAATCCAGAC 972  
Db 589 CTCGAAGGCCATTCCCTTGGCACAGACAGTGTTTTACGTGCTTTTACATTTCATAATCCCGAC 530  
QY 973 TTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTCTGGGAATCTTGAAAATCTGTGACATT 1032  
Db 529 TTCATAGAAGATCCTGCCATGAAAGCTTTTGTCTCTGGGAATCTTGAAGATCTGCGACATT 470  
QY 1033 GCAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTCAATGACT 1092  
Db 469 GCACGGGAAAAAGTAAATAAAGCTGCTGTTTTTCGAAGAGGAAGATTTCCAGTCAATGACA 410  
QY 1093 TATGGATTTAAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGGCATGCTAAAAGAT 1152  
Db 409 TACGGATTTAAAATGGCCAAACAGTGTGACAGATCTCCGAGTTACAGGGATGCTAAAAGAT 350  
QY 1153 GTGGAGGATGACATGCAAGAAGAGTAAAGAGTACTTCGAAGTCGACAAGGAGAGAAAGA 1212  
Db 349 GTGGAAGATGATCTGCAAAGCGAGTAAAGAGTACTTCGAAGTCGACAAGGAGAGAGAGA 290  
QY 1213 GATCCAGAAGTTGAACCTAGAACACCAACAATGTTTAGCAGTATTCAGCAGAGTGAAATTT 1272  
Db 289 GATCCAGAGGTCGAACCTAGAACCCACGAGTGTCTTGGCAGCATTTCAGCAGAGTGAAGTTC 230  
QY 1273 ACTCGTGTGTTACTGACAGTGTCTTATAGCCTTTTACTAGAAAAGAGACCAGTGTGTTGCA 1332  
Db 229 ACGGAGTGTGCTCACAGTGTCTATAGCCTTTTACTAAGAAAAGAGACCAGTGTGTTGCA 170  
QY 1333 GAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATTTCATAATTTCATTG 1392  
Db 169 GAGGCTCAGAAACTGATGGTCCAGCGGCAGACCTTCTTCTGCCATTTCACACCTCATTG 110  
QY 1393 CATCATGGCATCCAGGCCCAAGAATGATACTACAAAAGGAGATCATCCAATTATGATGGT 1452  
Db 109 CACCACGGCATCAAGGCTCCGAATGGCAGTACTACCAAAGGAGAGACCATCCAATTTAAATGGGT 50  
QY 1453 TTTGAACCCCTTGTGTAACCGAGGCTACTTCCACCTACCTTCCCTCGAT 1501  
Db 49 TTTGAGCCCTTGTGTTAACCGAGAGATTACTTCCCCCCCCGCTTCCCTCGNT 1

RESULT 10  
BM979511/c  
LOCUS  
DEFINITION  
UI-CF-DU1-adr-h-18-0-UI.s1 UI-CF-DU1 Homo sapiens cDNA clone  
UI-CF-DU1-adr-h-18-0-UI 3', mRNA sequence.  
ACCESSION  
BM979511  
VERSION  
BM979511.1 GI:19600035  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 616)  
AUTHORS  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE  
Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL  
Genome Res. 6 (9), 791-806 (1996)  
MEDLINE  
97044477  
PUBMED  
8889548  
COMMENT  
Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research

BM979511 616 bp mRNA linear EST 21-FEB-2003  
UI-CF-DU1-adr-h-18-0-UI.s1 UI-CF-DU1 Homo sapiens cDNA clone  
UI-CF-DU1-adr-h-18-0-UI 3', mRNA sequence.

Genetics ([www.resgen.com](http://www.resgen.com)) or from Open Biosystems  
([www.openbiosystems.com](http://www.openbiosystems.com)).  
Seq primer: M13 FORWARD  
POLYA=Yes.

```

FEATURES
source
Location/Qualifiers
1. .616
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-DUI1-adr-h-18-0-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-DUI1"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-CF-DUI1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Epithelial Cells The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT7T3-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GGCTGTAGGC.
TAG LIB=UI-CF-DUI1
TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
TAG_SEQ=GGCTGTAGGC"

```

```

Query Match          19.7%;   Score 610;   DB 12;   Length 616;
Best Local Similarity 100.0%;   Pred. No. 9.6e-100;
Matches 610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
BASE COUNT          190 a      123 c      98 g      205 t
ORIGIN

```

2487	QY	AGTTCAAGGAAATGTCTGACCTCAATAATAATAGCCCTCCTCCTCAGTCTCCTGAACTGT	2546
616	Db	AGTTCAAGGAAATGTCTGACCTCAATAATAATAGCCCTCCTCCTCAGTCTCCTGAACTGT	557
2547	QY	ATGTGGCAGCTAGTAAGCACTTTCAACAGGCAGCAAAATGATATTGGAAAAATATTCCTAAACC	2606
556	Db	ATGTGGCAGCTAGTAAGCACTTTCAACAGGCAGCAAAATGATATTGGAAAAATATTCCTAAACC	497
2607	QY	CGGACCATGAGGTTAATAGAAATTTTAAAGGTTGCCAAAAACCACACTTTGTGGTTATGAAAGT	2666
496	Db	CGGACCATGAGGTTAATAGAAATTTTAAAGGTTGCCAAAAACCACACTTTGTGGTTATGAAAGT	437
2667	QY	TATTGCCAGGAGGACACAAAAAGGAATCTAAAGTTTCCTCCTGAAATTTGATTTCTCTGCTC	2726
436	Db	TATTGCCAGGAGGACACAAAAAGGAATCTAAAGTTTCCTCCTGAAATTTGATTTCTCTGCTC	377
2727	QY	ATAAATAATTTTCCTGTTGTGAAACTTGTGAGAGAGACTGGGAGGTGGCCATAAAGGG	2786
376	Db	ATAAATAATTTTCCTGTTGTGAAACTTGTGAGAGAGACTGGGAGGTGGCCATAAAGGG	317
2787	QY	GCAGAGTCTTCTTTCCAGACCCCAACTCTTAGAGGGCACATCACCAGGCTCCACATCACGGG	2846
316	Db	GCAGAGTCTTCTTTCCAGACCCCAACTCTTAGAGGGCACATCACCAGGCTCCACATCACGGG	257
2847	QY	AAGTGAGATGGATTTCTTGGGTAACAACTCATTATAAGGAATACTTTTAGTTTGACAGCC	2906
256	Db	AAGTGAGATGGATTTCTTGGGTAACAACTCATTATAAGGAATACTTTTAGTTTGACAGCC	197
2907	QY	TTATATGACATGAATGAAAACTGCTGTTTTTAAAGTGGTTTTATTATGTTCCATGGAAGAAA	2966
196	Db	TTATATGACATGAATGAAAACTGCTGTTTTTAAAGTGGTTTTATTATGTTCCATGGAAGAAA	137
2967	QY	CTGGTCTTATTGAAATGCATTGATGAACGTTTATATGTTTTTATGTTTTTATACAGATTTTATCAAAA	3026
136	Db	CTGGTCTTATTGAAATGCATTGATGAACGTTTATATGTTTTTATGTTTTTATACAGATTTTATCAAAA	77

Qy	3027 TCATTTTTTATGAATGATTGAGTGAAATAAGTAGTTGTATATAAAGGTTAATAAATTTCTTGAC
Db	76 TCATTTTTTATGAATGATTGAGTGAAATAAGTAGTTGTATATAAAGGTTAATAAATTTCTTGAC 17
QY	3087 AAAAAAAAAA 3096       16 AAAAAAAAAA 7
Db	
RESULT 11	CA977646 LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS TITLE JOURNAL COMMENT
LOCUS	CA977646 AGENCOURT_11295472 NIH_MGC_164 Mus musculus cDNA clone IMAGE:30146905 5', mRNA sequence. CA977646 EST.
DEFINITION	CA977646.1 GI:27510300 Mus musculus (house mouse) EST.
ACCESSION	Mus musculus
VERSION	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
KEYWORDS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
SOURCE	1 (bases 1 to 900)
ORGANISM	NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC)
REFERENCE	Unpublished
AUTHORS	Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov
TITLE	Tissue Procurement: Dr. David Rowe and Dr. Mina
JOURNAL	CDNA Library Preparation: Invitrogen Corp
COMMENT	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: NDAM0063 row: f column: 02 High quality sequence stop: .655. Location/Qualifiers 1. .900 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:30146905" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_164" /note="Vector: pCMV-SPORT6.1.ccd; Site_1: EcoRV; Site_2: NotI; Non-normalized full-length enriched library from pooled mouse embryonic limb, maxilla and mandible, day 10.5 and 11.5 (size selected for the 0.5-1 kb fragments) Cloned directionally, priming method: Oligo-dt. CDNA enrichment: >1k bp, Average insert size 1.8k bp. Priming sequence: 5'GACTAGTTCTAGATCGCAGCGGCCCC(T) 3'. Tissue contributed by, David Rowe. Library constructed by ResGen, Invitrogen Corp."
FEATURES	source
BASE COUNT	235 a   213 c   212 g   239 t       1 others
ORIGIN	

[illegible]



DEFINITION wt02f11.x1 NCI\_CGAP\_Co3 Homo sapiens cDNA clone IMAGE:2506317 3',  
mRNA sequence.  
ACCESSION AW007583  
VERSION AW007583.1 GI:5856446  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 598)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arraying: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 715 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 459.  
Location/Qualifiers  
1. .598  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2506317"  
/sex="pooled"  
/tissue\_type="colon"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP Co3"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from 12 pooled bulk tumor samples and primed  
with a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT7T3 vector. Library went through one round of  
normalization."   
189 a 122 c 98 g 188 t 1 others  
BASE COUNT  
ORIGIN  
Query Match 19.3%; Score 597; DB 9; Length 598;  
Best Local Similarity 99.8%; Pred. No. 2.1e-97;  
Matches 597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2490 TCAAGGAATGTCTGACCTCAATAATAATATAGCCCTCCTCAGTCTCCTGAACTGTATG 2549  
Db |||||  
QY 2550 TGGCAGCTAGTAAGCACCTTCAACAGGCAAAAAATGATATTGGAATAATTCCTAACCCGG 2609  
Db |||||  
QY 538 TGGCAGCTAGTAAGCACCTTCAACAGGCAAAAAATGATATTGGAATAATTCCTAACCCGG 479  
QY 2610 ACCATGAGGTTAATAGAATTTTAAAGGTTGCCAAACCCCACTTTTGGTTATGAAGTTAT 2669  
Db |||||  
QY 478 ACCATGAGGTTAATAGAATTTTAAAGGTTGCCAAACCCCACTTTTGGTTATGAAGTTAT 419  
QY 2670 TGGCAGGAGGACACAAAAGGAATCTAAAGTTCCTCCTGAATTTGATTTCTCTGCTCAT 2729  
Db |||||  
QY 418 TGGCAGGAGGACACAAAAGGAATCTAAAGTTCCTCCTGAATTTGATTTCTCTGCTCAT 359  
QY 2730 AATATTTTCTGTTGTGAAACTTGTGAGAGAGACTGGGAGGTGGCCATAAAGGGCA 2789  
Db |||||  
QY 358 AATATTTTCTGTTGTGAAACTTGTGAGAGAGACTGGGAGGTGGCCATAAAGGGCA 299  
QY 2790 GAGTCTTCTTTCAGACCCCAACTCTTAGAGGGACATCACCAGGCTCCACATCACGGGAAG 2849

Db 298 GAGTCTTCTTTCAGACCCCAACTCTTAGAGGGCACATCACCAGGCTCCACATCACGGGAAG 239  
QY 2850 TGAGATGGATTTCTTGGGTAACAACTCATTATAAGGATACTTTTAGTTTGACAGCCTTA 2909  
Db 238 TGAGATGGATTTCTTGGGTAACAACTCATTATAAGGATACTTTTAGTTTGACAGCCTTA 179  
QY 2910 TATGACATGAATGAAACTGCTGTTTTTAAAGTGGTTTATATGTTCCATGGAAGAACTG 2969  
Db 178 TATGACATGAATGAAACTGCTGTTTTTAAAGTGGTTTATATGTTCCATGGAAGAACTG 119  
QY 2970 GTCTTATTGAATGCATTGATGAACGTTATATATGTTTATACAGATTTAATCACAATCA 3029  
Db 118 GTCTTATTGAATGCATTGATGAACGTTATATGTTTATTACAGATTTAATCACAATCA 59  
QY 3030 TTTTATTGAATGATTGAGTGAAAAATAGTGTTTATAAAGGTTAATAAAATTTCTTGACA 3087  
Db 58 TTTTATTGAATGATTGAGTGAAAAATAGTGTTTATAAAGGTTAATAAAATTTCTTGACA 1  
RESULT 14  
BU058931  
LOCUS UI-M-FR0-cam-e-22-0-UI.r1 NIH\_BMAP\_FR0 Mus musculus cDNA clone  
DEFINITION IMAGE:6414333 5', mRNA sequence.  
ACCESSION BU058931  
VERSION BU058931.1 GI:22499220  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 718)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)  
Seq primer: pYX-5.  
Location/Qualifiers  
1. .718  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6414333"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH\_BMAP\_FR0"  
/note="Organ: Brain; Vector: pYX- Asc; Site\_1: EcoR I;  
Site\_2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is AGCGAGACAG. This library was created for the University  
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,

```

BASE COUNT      222 a   159 c   165 g   171 t       1 others
ORIGIN

Query Match      19.2%;   Score 593.8;   DB 13;   Length 718;
Best Local Similarity 89.1%;   Pred. No. 7.8e-97;
Matches 640;   Conservative 0;   Mismatches 78;   Indels 0;   Gaps 0;

QY      896  GCTGTTTGATAACGTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTACGTGCCTTT 955
Db      1    GCTGTTTGATCACATGGCTCGAAGGCCATTCTTGGCACAGACAGTGTTCAGTGCCCTTT 60

QY      956  ACATTCATAATCCAGACTTTATAGAAAGATCCTGCTATGAAGCCTTTTGCTCTGSGAATCT 1015
Db      61    ACATTCATAATCCCGACTTCATAGAAAGATCCTGCCATGAAAGCTTTTGCTCTGGGAATCT 120

QY      1016  TGAATAATCTGTGACATTGCAAGGGAAGAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAG 1075
Db      121  TGAAGATCTGCGACATTGCAACGGGAAGAAAGTAAATAAAGCTGCTGTTTTCGAAGAGGAAG 180

QY      1076  ATTTTCAGTCAATGACTTATGGATTAAATGGCTAACAGTGTGACAGATCTTCGAGTTA 1135
Db      181  ATTTCCAGTCAATGACATACGGAATTAAATGGCCAAACAGTGTGACAGATCTCCGAGTTA 240

QY      1136  CAGGCATGCTAAAGATGTGGAGGATGACATGCAAAGAGAGATAAGAGTACTCGAAGTC 1195
Db      241  CAGGGATGCTAAAGGATGTGGAAGATGATCTGCAAAAGGCGAGTAAAGAGTACTCGAAGTC 300

QY      1196  GACAAAGGAGAAAGAGATCAGAAAGTTGAACTAGAACACCAACAATGTTTAGCAGTAT 1255
Db      301  GACAAAGGAGAAAGAGAGATCCAGAGGTCGAACTAGAACACACAGCAGTCTTGGCAGCAT 360

QY      1256  TCAGCAGAGTGAAATTTACTCGTGTGTACTGACAGTGTCTATAGCCTTTACTAAGAAAG 1315
Db      361  TCAGCAGAGTGAAGTTCACGGGAGTGTGCTCACAGTGTCTATAGCCTTTACTAAGAAAG 420

QY      1316  AGACCAGTGTGTTGCAGAAAGCTCAAAAATTGATGTTCAAGCAGCAGATCTTCTTTCTG 1375
Db      421  AGACCAGTGTGTTGCAGAGGCTCAGAAACTGATGGTCCAGGGCGCAGACCTTCTTTCTG 480

QY      1376  CCATTCATAATTCAATGTCATCATGGCATCCAGGCCCAGAAATGATACATAAAGGAGATC 1435
Db      481  CCATTACACCTCATTTGCACCACGGCATCCAGGCTCAGAAATGGCACTACCAAGGAGACC 540

QY      1436  ATCCAATTATGATGGGTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCC 1495
Db      541  ATCCAATTATGATGGGTTTTGAGCCCTTGTGTAACCCAGAGATTACTTCCACCCACCTTCC 600

QY      1496  CTCGATATGCAAAAATAATTAAAGGGAAGAAATGGTGAACATAATTTGCAAGATTAAATAG 1555
Db      601  CTCGCTATGCNAAAATAATTAAAGAGAGAAATGGTCAAACTATTCTCAAGATTAAATAG 660

QY      1556  ATAGAATAAAAACTGCTGTGAGGTTGTGAATTTAACAATTTTACATTGTATCCTGGA 1613
Db      661  ACAGAATAAAAACTGCTGTGAAGTCGTGAACCTTACCAAACTTACACTGTATCCTGGA 718

RESULT 15
BG112812
LOCUS
DEFINITION 602285005F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4372320 5',
           mRNA sequence.
ACCESSION BG112812
VERSION   BG112812.1 GI:12606318
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 972)
AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished
```

```

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10032 row: j column: 01
High quality sequence stop: 658.
Location/Qualifiers
1..972
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4372320"
/tissue_type="osteosarcoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_86"
/note="Organ: bone; Vector: pCMV-SPORT6; Site 1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.533 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT      285 a   232 c   214 g   241 t
ORIGIN

Query Match      18.9%;   Score 586;   DB 10;   Length 972;
Best Local Similarity 96.6%;   Pred. No. 1.9e-95;
Matches 663;   Conservative 0;   Mismatches 15;   Indels 8;   Gaps 6;

QY      2090  CTTCTAAGTGGCTTTGAAATTGGAACTCTACAGTATGCACGAGTACTATTACATATATTGG 2149
Db      3    CTTCTAAGTGGCTTTGAA-TGGAACTCTACAGTATGCACGAGTACTATTACATATA-TGG 60

QY      2150  TATCTCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTGAGTCGTCGCCGATGGCTCT 2209
Db      61  TATCTCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTGAGTCGTCGCCGATGGCTCT 120

QY      2210  CAAATGGCAGAGGAAAGGATAATGGAAGAGCAGCAGAAAGCCGCTAGTAGTAAAAAACA 2269
Db      121  CAAATGGCAGAGGAAAGGATAATGGAAGAGCAGCAGAAAGCCGCTAGTAGTAAAAAACA 180

QY      2270  AAGAAAAAAGAAAGTTCCGCCATTGAGCCGAGAGATCAAAATGAGCCCAAGCATATCAG 2329
Db      181  AAGAAAAAAGAAAGTTCCGCCATTGAGCCGAGAGATCAAAATGAGCCCAAGCATATCAG 240

QY      2330  AACATGTGTCTGGAAATGTTTAAACCATTGGTAGCATTTGACATGGACGGCAAGTACGT 2389
Db      241  AACATGTGTCTGGAAATGTTTAAACCATTGGTAGCATTTGACATGGACGGCAAGTACGT 300

QY      2390  AAACCGAAGTTTGAGCTTGATAGTGAACAAGTTCGGTATGAACACAGGTTTGTCTCCATT 2449
Db      301  AAACCGAAGTTTGAGCTTGATAGTGAACAAGTTCGGTATGAACACAGGTTTGTCTCCATT 360

QY      2450  AACAGTGTGATGACCCCGCCGCGCAGTGCACCTACTTACAGTTCAAGGAAATGTCTGACCT 2509
Db      361  AACAGTGTGATGACCCCGCCGCGCAGTGCACCTACTTACAGTTCAAGGAAATGTCTGACCT 420

QY      2510  AATAAATATAGCCCTCCTCCTCAGTCTCCTGAACTGATGTGGCAGCTAGTAAGCACTTT 2569
Db      421  AATAAATATAGCCCTCCTCCTCAGTCTCCTGAACTGATGTGGCAGCTAGTAAGCACTTT 480

QY      2570  CAACAGGCAAAAATGATATTGGAAAAATATTCCTAACCCGGACCATTGAGTTAATAGAATT 2629
Db      481  CAACAGGCAAAAATGATATTGGAAAAATATTCCTAACCCGGACCATTGAGTTAATAGAATT 540

QY      2630  TTAAAGGTTGCCAAACCCAACTTTGTGGTT-ATGAAGTTATTGGCAGGAGACACAA-AA 2687
Db      541  TTAAAGGTTGCCAAACCCAACTTTGTGGTTTCATGAAGTTATTGGCAGGAGACACAA 600

QY      2688  AGGAATCTAAAGTTCCTCTGAATTTGATTTCTCTCTCATATAAATATTTTCTGTGTGA 2747
```

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10032 row: j column: 01  
High quality sequence stop: 658.

FEATURES  
source

1..972  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4372320"  
/tissue\_type="osteosarcoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_86"  
/note="Organ: bone; Vector: pCMV-SPORT6; Site 1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.533 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

```
Db      601 AGGAATCTAAAGTTCTCTCCTGAA-TTGATTTCCTGCTCATA---ATATTTCCTGTGTGA 656
Qy      2748 AAC TTGTTTGAGAGAGACTGGGGAGG 2773
Db      657 AAC TGTGAGAGAGACTGGGAGGTGG 682
```

Search completed: November 25, 2003, 02:50:11  
Job time : 6201 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 24, 2003, 22:01:10 ; Search time 96.5 Seconds  
(without alignments)  
10184.818 Million cell updates/sec

Title: US-10-001-857-42  
Perfect score: 5439  
Sequence: 1 ttcttcacgaactccagg.....atttcttgacaaaaaaaaa 3096

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgn2\_1/uspt0/spool/us10001857/runat\_24112003\_162854\_9527/app\_query.fasta\_1.3271  
-DB=A\_Geneseq\_19Jun03 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10001857 @CGN\_1\_1\_0 @runat\_24112003\_162854\_9527 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_19Jun03:  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3796	69.8	733	23	ABP52902	Human lung specifi
2	3732.5	68.6	725	23	AAO19400	Human molecule for
3	1294.5	23.9	784	22	ABB60388	Drosophila melanog
4	1237	22.7	239	23	ABG69816	Human secretory pr
5	392	7.1	108	22	AAU32414	Novel human secret
6	158.5	2.9	681	23	ABB93650	Herbicidally activ
7	153.5	2.8	156	22	ABG14149	Novel human diagno
8	153	2.8	708	23	ABB91504	Herbicidally activ
9	146.5	2.7	120	22	ABU52794	Human transmembran
10	146.5	2.7	574	22	AAU52322	WASP homolog prote
11	146.5	2.7	574	22	AAG67370	Amino acid sequenc
12	146	2.7	416	23	ABG79529	Human tumour suppr
13	145.5	2.6	694	23	ABB92421	Herbicidally activ
14	145	2.6	149	22	AAO05346	Human polypeptide
15	145	2.6	559	22	AAU52317	Human Scari protei
16	145	2.6	559	22	AAG67345	Amino acid sequenc
17	145	2.6	559	22	AAU20494	Human Pablo (pro-a
18	145	2.6	559	22	AAU73399	Human WAVE1 (WAVE/
19	144	2.6	1134	23	AAE24341	Human lung-specifi
20	143.5	2.6	19938	24	ABP76682	Streptomyces virid
21	143	2.6	598	22	ABG14000	Novel human diagno
22	143	2.6	780	24	ABU00297	Human novel polype
23	143	2.6	1248	20	AAU13464	Human diaphanous p
24	143	2.6	19938	24	ABB98398	Streptomyces virid
25	141.5	2.6	551	23	ABG93105	S. cerevisiae BAX-
26	141.5	2.6	560	22	ABG21040	Novel human diagno
27	141.5	2.6	5877	24	AAE34702	Human mucin (MUC-1
28	141	2.6	191	21	AAG29836	Arabidopsis thalia
29	141	2.6	199	21	AAG29835	Arabidopsis thalia
30	140.5	2.6	19938	24	ABP76680	Streptomyces virid
31	140	2.5	541	19	AAW37148	Mammalian Ena (Men
32	140	2.5	783	19	AAW37151	Mouse neural Mena+
33	140	2.5	787	19	AAW37152	Mouse neural Mena+
34	140	2.5	802	19	AAW37153	Mouse neural Mena+
35	140	2.5	802	22	AAU09139	Mammalian enabled
36	139.5	2.5	1243	24	AAE34865	Human kinases and
37	139.5	2.5	1243	24	ABP71620	Human WNK4 protein
38	139	2.5	214	17	AAU86913	Cotton fiber-speci
39	138.5	2.5	250	21	AAU67470	Np70 protein carbo
40	138.5	2.5	641	21	AAU82327	Human Npw38BP1 tra
41	138.5	2.5	641	21	AAU67469	Np70 protein seque
42	138.5	2.5	641	22	AAU72165	Human RNA metaboli
43	138.5	2.5	641	22	AAU47514	NpwBP, Homo sapie
44	138.5	2.5	647	21	AAU53462	Human colon cancer
45	138.5	2.5	903	23	ABP65235	Hypoxia-regulated

ALIGNMENTS

RESULT 1  
ABP52902  
ID ABP52902 standard; Protein; 733 AA.  
XX  
AC ABP52902;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE Human lung specific protein sequence SEQ ID NO:145.  
XX  
KW Human; lung; lung specific nucleic acid; LSNA; lung specific protein;  
KW LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;  
KW squamous cell carcinoma.  
XX  
OS Homo sapiens.  
XX  
PN WO200264788-A2.  
XX

PD 22-AUG-2002.  
XX  
PF 20-NOV-2001; 2001WO-US45080.  
XX  
PR 20-NOV-2000; 2000US-252054P.  
XX  
PA (DIAD-) DIADEXUS INC.  
XX  
PI Macina RA, Recipon H, Chen S, Sun Y, Liu C;  
XX  
DR WPI; 2002-657601/70.  
XX  
PT New lung specific nucleic acid useful in gene therapy or as vaccines  
PT for treating lung cancer (e.g. squamous cell carcinoma) or  
PT non-cancerous lung diseases, as well as for diagnosing, monitoring or  
PT staging these diseases -  
XX  
PS Claim 11; Page 239-242; 282pp; English.  
XX

CC The present invention describes an isolated lung specific nucleic acid  
(LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid  
CC sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;  
CC (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp),  
CC given in ABQ75262 to ABQ75376; (c) selectively hybridizes to (a) or (b);  
CC or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific  
CC protein (LSP) sequences have cytostatic activity and can be used in gene  
CC therapy and vaccines. LSNA and LSPs are useful for diagnosing and  
CC monitoring the presence and metastases of lung cancer in a patient. An  
CC antibody that specifically binds to an LSP can be used for determining  
CC the presence of an LSP in a sample, as well as for treating a patient  
CC with lung cancer, particularly by inducing an immune response against  
CC the lung cancer cell expressing the LSNA or LSPs. In particular, these  
CC LSNA and LSPs are useful for identifying, diagnosing, monitoring,  
CC staging, imaging and treating lung cancer (e.g. squamous cell carcinoma)  
CC and non-cancerous disease states in lung.

XX  
SQ Sequence 733 AA;

Alignment Scores:

Pred. No.: 0 Length: 733  
Score: 3796.00 Matches: 732  
Percent Similarity: 99.59% Conservative: 0  
Best Local Similarity: 99.59% Mismatches: 1  
Query Match: 69.79% Indels: 3  
DB: 23 Gaps: 0

US-10-001-857-42 (1-3096) x ABP52902 (1-733)

QY	556	ATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCTCAGTATGCCAGAA	615
Db	1	MetValMetLysAlaSerValAspAspAspSerGlyTrpGluLeuSerMetProGlu	20
QY	616	AAATGAGAAAAGCAATACAACTGGGTGGACATTACCCAGATTTTGAAGAAGCTTGT	675
Db	21	LysMetGluLysSerAsnThrAsnTrpValAspIleThrGlnAspPheGluGluAlaCys	40
QY	676	CGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCTATTTGGTCTTTTGAAGCCATG	735
Db	41	ArgGluLeuLysLeuGlyGluLeuLeuHisAspLysLeuPheGlyLeuPheGluAlaMet	60
QY	736	TCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAAACCAAGTTAAT	795
Db	61	SerAlaIleGluMetMetAspProLysMetAspAlaGlyMetIleGlyAsnGlnValAsn	80
QY	796	CGAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAATTAAGATCTC	855
Db	81	ArgLysValLeuAsnPheGluGlnAlaIleLysAspGlyThrIleLysIleLysAspLeu	100
QY	856	ACCTTGCTGAACCTGATAGGATTATGGATACATGTTTTTGGTGTGATAACGTGGTTA	915
Db	101	ThrLeuProGluLeuIleGlyIleMetAspThrCysPheCysCysLeuIleThrTrpLeu	120
QY	916	GAAGGCCATTCACTGGCACAGACAGATTAATACGTGCCTTTTACATTCATAATCCAGACTT	975

Db	121	GluGlyHisSerLeuAlaGlnThrValPheThrCysLeuTyrlleHisAsnProAspPhe	140
QY	976	ATAGAAGATCCTGCTATGAAGGCTTTTGGCTCTGGAATCTTTGAAAATCTGTGACATTGCA	1035
Db	141	IleGluAspProAlaMetLysAlaPheAlaLeuGlyIleLeuLysIleCysAspIleAla	160
QY	1036	AGGAAAAAGTAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTCAATGACTTAT	1095
Db	161	ArgGluLysValAsnLysAlaAlaValPheGluGluAspPheGlnSerMetThrTyr	180
QY	1096	GGATTTAAAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGCATGCTAAAAAGATGTG	1155
Db	181	GlyPheLysMetAlaAsnSerValThrAspLeuArgValThrGlyMetLeuLysAspVal	200
QY	1156	GAGGATGACATGCAAGAAGAGTAAAGAGTACTCGAAGTCGACAAGGAGAGAAAGAGAT	1215
Db	201	GluAspAspMetGlnArgArgValLysSerThrArgSerArgGlnGlyGluGluArgAsp	220
QY	1216	CCAGAAAGTTGAACACCAACCAATGTTTAGCAGTATTTCAGCAGAGTGAATTTACT	1275
Db	221	ProGluValGluLeuGluHisGlnGlnCysLeuAlaValPheSerArgValLysPheThr	240
QY	1276	CGTGTGTTACTGACAGTGTATAGCCTTTACTAAGAAAGAGACCAGTGTCTGTTGCAGAA	1335
Db	241	ArgValLeuLeuThrValleuIleAlaPheThrLysLysGluThrSerAlaValAlaGlu	260
QY	1336	GCTCAAAAATTTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATTTCATAATTCATTGCAT	1395
Db	261	AlaGlnLysLeuMetValGlnAlaAlaAspLeuLeuSerAlaIleHisAsnSerLeuHis	280
QY	1396	CATGGCATCCAGGCCAGATGATATACAAAAGGAGATCATCCAATTATGATGGGTTT	1455
Db	281	HisGlyIleGlnAlaGlnAsnAspThrThrLysGlyAspHisProIleMetMetGlyPhe	300
QY	1456	GAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACTTCCCTCCGATATGCAAAAATAAT	1515
Db	301	GluProLeuValAsnGlnArgLeuLeuProProThrPheProArgTyrAlaLysIleIle	320
QY	1516	AAAAGGGAAGAAATGGTGAACCTATTTTTGCAAGATTAAATAGATAGATAAAACCTGTCTGT	1575
Db	321	LysArgGluGluMetValAsnTyrPheAlaArgLeuIleAspArgIleLysThrValCys	340
QY	1576	GAGTTGTGAATTTAACAATTTACATTGTATCCTGGATTTTCTGTGAATTTAGTGAA	1635
Db	341	GluValValAsnLeuThrAsnLeuHisCysIleLeuAspPheCysGluPheSerGlu	360
QY	1636	CAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTCTCTGGTGGATAACAAA	1695
Db	361	GlnSerProCysValLeuSerArgSerLeuLeuGlnThrThrPheLeuValAspAsnLys	380
QY	1696	AAGTCTTTTGAACCTCATCTCATGCAAGACATGGTGAAAGATGCACTTTCGGTCTTTTGTCT	1755
Db	381	LysValPheGlyThrHisLeuMetGlnAspMetValLysAspAlaLeuArgSerPheVal	400
QY	1756	AGATCCTCCGAGTGTCTTCCCCAAGTGCTACCTATATAATAATCACCAGGCTAAGGACT	1815
Db	401	-AspProPro-ValLeuSerProLysCysTyrLeuTyAsnAsnHisGlnAlaLysAspC	420
QY	1816	GTATCGACTCCTTTGTTACTCAGTGTGTTCCGCCATTCTCTGAGTCTTATTTCAGATCCATG	1875
Db	420	ysIleAspSerPheValThrHisCysValArgProPheCysSerLeuIleGlnIleHisG	440
QY	1876	GACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAATTTGCCACCT	1935
Db	440	lyHisAsnArgAlaArgGlnArgAspLysLeuGlyHisIleLeuGluGluPheAlaThrL	460
QY	1936	TGACGATGAGTTTATGACATTTTATTTTAAATAGGCA-GAGAAGGTTGATGCAGCGCTTC	1994
Db	460	euGlnAspGluPheMetThrPheTyrPheAsnArgAlaGluLysValAspAlaAlaLeuH	480
QY	1995	ACACCATGCTGTTGAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGGTACCTGGG	2054

Db 480 isThrMetLeuLeuLysGlnGluProGlnArgGlnHisLeuAlaCysLeuGlyThrTrpV 500

QY 2055 TCCTTTACCATAAACCTTCGCATTATGATACAGTACTTCTAAGTGGCTTTGAATTGGAAC 2114

Db 500 alLeuTyrHisAsnLeuArgIleMetIleGlnTyrLeuLeuSerGlyPheGluLeuGluL 520

QY 2115 TCTACAGTATGCACGACTACTATTACATATATTGGTATCTCTCTGAATTCCTTTACGCAT 2174

Db 520 euTyrSerMetHisGluTyrTyrTyrIleTyrTrpTyrLeuSerGluPheLeuTyrAlaT 540

QY 2175 GGTGATGTCAACATTGAGTCGTGCCGATGGCTCTCAAAATGGCAGAGAAAGGATAATGG 2234

Db 540 rpLeuMetSerThrLeuSerArgAlaAspGlySerGlnMetAlaGluGluArgIleMetG 560

QY 2235 AAGAGCAGCAGAAAGGCCGTAGTAGTAAATAAACAAGAAAAAAGAAAGTTTCGCCCAT 2294

Db 560 luGluGlnGlnLysGlyArgSerSerLysLysThrLysLysLysLysLysValArgProL 580

QY 2295 TGAGCCGAGAGATCACAATGAGCCCAAGCATATCAGAACATGTGTGCTGGAATGTTAAAA 2354

Db 580 euSerArgGluIleThrMetSerGlnAlaTyrGlnAsnMetCysAlaGlyMetPheLysT 600

QY 2355 CCATGGTAGCATTTGACATGGACGGCAAAGTACGTAAACCGAAGTTTGAGCTTGATAGTG 2414

Db 600 hrMetValAlaPheAspMetAspGlyLysValArgLysProLysPheGluLeuAspSerG 620

QY 2415 AACAAAGTTCGGTATGAACACAGGTTTGTCTCCATTCAACAGTGTGATGACCCCGCGCCAG 2474

Db 620 luGlnValArgTyrGluHisArgPheAlaProPheAsnSerValMetThrProProProV 640

QY 2475 TGCACACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCCCTCCTCAGT 2534

Db 640 alHisTyrLeuGlnPheLysGluMetSerAspLeuAsnLysTyrSerProProProGlnS 660

QY 2535 CTCCTGAACGTGTATGGCAGCTAGTAGAACACACTTTCAACAGGCAAAAATGATATTGGAAA 2594

Db 660 erProGluLeuTyrValAlaAlaSerLysHisPheGlnGlnAlaLysMetIleLeuGluA 680

QY 2595 ATATTCTTAACCCGGACCATGAGGTTAATAGAAATTTAAAGGTTGCCAAACCCAACTTTG 2654

Db 680 snIleProAsnProAspHisGluValAsnArgIleLeuLysValAlaLysProAsnPheV 700

QY 2655 TGGTTATGAAGTTATTGGCAGGAGGACACAAAAGGAATCTAAAGTTCCTCCTCAATTTG 2714

Db 700 alValMetLysLeuLeuAlaGlyGlyHisLysLysGluSerLysValProProGluPheA 720

QY 2715 ATTTCTCTGCTCATAAATATTTTCTCTGTTGTGAAACTTGTT 2755

Db 720 spPheSerAlaHisLysTyrPheProValValLysLeuVal 733

RESULT 2

AAO19400

ID AAO19400 standard; Protein; 725 AA.

XX AAO19400;

AC AAO19400;

XX 10-DEC-2002 (first entry)

DT Human molecule for disease detection and treatment protein #3.

XX Human; molecule for disease detection and treatment; MDDT; gene therapy;

KW cytotstatic; antiarteriosclerotic; hepatotropic; anti-HIV; antiallergic;

KW antiinflammatory; antiasthmatic; cerebroprotective; nootropic;

KW neuroprotective; antiparkinsonian; cardiant; antianginal.

XX Homo sapiens.

OS WO200270709-A2.

XX 12-SEP-2002.

PD 08-FEB-2002; 2002WO-US03709.

XX

PR 09-FEB-2001; 2001US-268117P.

PR 15-FEB-2001; 2001US-269618P.

PR 23-FEB-2001; 2001US-271118P.

PR 07-MAR-2001; 2001US-274486P.

PR 09-MAR-2001; 2001US-274436P.

PR 28-NOV-2001; 2001US-334229P.

PR 01-FEB-2002; 2002US-353284P.

XX (INCY-) INCYTE GENOMICS INC.

PA Lal PG, Baughn MR, Yao MG, Walia NK, Elliot VS, Xu Y;

XX Honchell CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DAM;

PI Hafalia AJA, Ghandi AR, Thangavelu K, Sanjanwala MM, Tang YT;

PI Ramkumar J, Griffiin JA, Swarnaker A, Azimzai Y, Sapperstein SK;

PI Burford N, Lee EA, Lu Y, Tran UK, Marquis JP;

XX WPI; 2002-713453/77.

DR N-PSDB; AAL49929.

XX New human molecules for disease detection and treatment (MDDT), useful

PT for diagnosing, treating and preventing diseases or conditions

PT associated with the aberrant MDDT expression, e.g. cancer, AIDS,

PT asthma, diabetes, hepatitis -

XX Claim 1; Page 128-130; 177pp; English.

PS The present invention relates to human proteins and coding sequences of

XX molecules for disease detection and treatment MDDT. The sequences can be

CC used in the treatment of diseases associated with the decreased

CC expression or overexpression of MDDT, such as cell proliferative (cancer,

CC atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS,

CC allergies, Addison's disease, asthma), developmental (dwarfism, renal

CC tubular acidosis), neurological (e.g. stroke, Parkinson's disease,

CC epilepsy) and cardiovascular (congestive heart failure, myocardial

CC infarction, angina pectoris) disorders. The present sequence is a protein

CC of the invention.

XX SQ Sequence 725 AA;

Alignment Scores:

Pred. No.: 0 Length: 725

Score: 3732.50 Matches: 723

Percent Similarity: 98.23% Conservative: 0

Best Local Similarity: 98.23% Mismatches: 2

Query Match: 68.62% Indels: 11

DB: 23 Gaps: 1

US-10-001-857-42 (1-3096) x AAO19400 (1-725)

QY 556 ATGTTATGAAAGCTTCTGTAGATGATGACGATTGAGGAGCTCAGTATGCCAGAA 615

Db 1 MetValMetLysAlaSerValAspAspAspSerGlyTrpGluLeuSerMetProGlu 20

QY 616 AAAATGGAGAAAAGCAATACAACTGGTGGACATTACCCAAAGATTTTGAAGAGCTTGT 675

Db 21 LysMetGluLysSerAsnThrAsnTrpValAspIleThrGlnAspPheGluGluAlaCys 40

QY 676 CGAGAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTTGGTCTTTTGAAGCCATG 735

Db 41 ArgGluLeuLysLeuGlyGluLeuLeuHisAspLysLeuPheGlyLeuPheGluAlaMet 60

QY 736 TCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAAACCAAGTTAAT 795

Db 61 SerAlaIleGluMetMetAspProLysMetAspAlaGlyMetIleGlyAsnGlnValAsn 80

QY 796 CGAAAAGTTCCTCAATTTTGAACAAGCTATCAAGGATGGCAGTATTTAAATTAAGATCTC 855

Db 81 ArgLysValLeuAsnPheGluGlnAlaIleLysAspGlyThrIleLysIleLysAspLeu 100

QY 856 ACCTTGCCCTGAAGTATAGGGATTATGGATACATGTTTGTCTGTTTGTGATAACGTGTTA 915

Db 101 ThrLeuProGluLeuIleGlyIleMetAspThrCysPheCysCysLeuIleThrTrpLeu 120

QY	916	GAAGGCCATTCACTGGCAGACAGATATTTACGTGCCTTTACATTATAATCCAGACTTT	975
Db	121	GluGlyHisSerLeuAlaGlnThrValPheThrCysLeuTyrIleHisAsnProAspPhe	140
QY	976	ATAGAAGATCCTGCTATGAAGCTTTTGCTCTGGGAATCTTGAAAATCTGTGACATTGCA	1035
Db	141	IleGluAspProAlaMetLysAlaPheAlaLeuGlyIleLeuLysIleCysAspIleAla	160
QY	1036	ACGGAAAAAGTAAAGCTGCTGTTTTTTGAAGAGGAAGATTTTCAGTCAATGACTTAT	1095
Db	161	ArgGluLysValAsnLysAlaAlaValPheGluGluGluAspPheGlnSerMetThrTyr	180
QY	1096	GGATTTAAATGGCTAACAGGTGCACAGATCTTCGAGTTACAGGCATGCTAAAAGATGTG	1155
Db	181	GlyPheLysMetAlaAsnSerValThrAspLeuArgValThrGlyMetLeuLysAspVal	200
QY	1156	GAGGATGACATGCAAAGAGTAAGAGTACTCGAAGTCGACAAAGGAGAGAAGAGAT	1215
Db	201	GluAspAspMetGlnArgArgValLysSerThrArgSerArgGlnGlyGluGluArgAsp	220
QY	1216	CCAGAAGTTGAAC TAGACACACCAACAATGTTTAGCAGTATTCAGCAGAGTGAAATTTACT	1275
Db	221	ProGluValGluLeuGluHisGlnGlnCysLeuAlaValPheSerArgValLysPheThr	240
QY	1276	CGTGTGTTACTGACAGTGCTTATAGCCTTTACTAAGAAAGAGACCAGTGCTGTTGCAGAA	1335
Db	241	ArgValLeuLeuThrValLeuIleAlaPheThrLysLysLysGluThrSerAlaValAlaGlu	260
QY	1336	GCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATTCATAATTCATTGCAT	1395
Db	261	AlaGlnLysLeuMetValGlnAlaAlaAspLeuLeuSerAlaIleHisAsnSerLeuHis	280
QY	1396	CATGSCATCCAGGCCCAGAAATGATACTACAAAAGGAGATCATCCAATATATGATGGTTTT	1455
Db	281	HisGlyIleGlnAlaGlnAsnAspThrThrLysGlyAspHisProIleMetMetGlyPhe	300
QY	1456	GAACCCCTTGTAACCAAGGCTACTTCCACCTACCTTCCCTCGATATGCAAAAATAATT	1515
Db	301	GluProLeuValAsnGlnArgLeuLeuProProThrPheProArgTyrAlaLysIleIle	320
QY	1516	AAAAGGAAGAAATGGTGAACTATTTTCGAAGATTAATAGATAGAATAAAAACGTCTGT	1575
Db	321	LysArgGluGluMetValAsnTyrPheAlaArgLeuIleAspArgIleLysThrValCys	340
QY	1576	GAGGTGTGAATTTAACAAATTTACATTTGATCTCCTGGATTTTCTGTGAATTTAGTGAA	1635
Db	341	GluValValAsnLeuThrAsnLeuHisCysIleLeuAspPhePheCysGluPheSerGlu	360
QY	1636	CAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACCTTTCCTGGTGGATAACAA	1695
Db	361	GlnSerProCysValLeuSerArgSerLeuLeuGlnThrThrPheLeuValAspAsnLys	380
QY	1696	AAGTCTTTGGAACTCATCTCATGCAAGATGGTGAAAGATGCATTCGGTCTTTTGTC	1755
Db	381	LysValPheGlyThrHisLeuMetGlnAspMetValLysAspAlaLeuArgSerPheVal	400
QY	1756	AGATCCTCCGAGTGCTTTCCCCCAAGTGCTACCTATATAATAATCACCAAGGCTAAGGACT	1815
Db	401	-SerProPro-ValLeuSerProLysCysTyrLeuTyrAsnAsnHisGlnAlaLysAspC	420
QY	1816	GTATCGACTCCTTTGTTACTCACTGTGTTCGGCCATTCTGTAGTCTTATTCAGATCCATG	1875
Db	420	ysIleAspSerPheValThrHisCysValArgProPheCysSerLeuIleGlnIleHisG	440
QY	1876	GACATAACAGGCTCGACAGAGAGATAAGTTGGTCATATTTCTTGAGGAATTTGCCACCT	1935
Db	440	lyHisAsnArgAlaArgGlnArgAspLysLeuGlyHisIleLeuGluGluPheAlaThrL	460
QY	1936	TGCAGGATGAGTTTATGACATTTTATTTTAATAGGCAGAGAAGGTTGATGCAGCGCTTCA	1995
Db	460	euGlnAsp-----Glu--AlaGluLysValAspAlaAlaLeuHi	472
QY	1996	CACCATGCTGTGAAAACAGGAACCCCAAGGCAACATTTGGCCTGTTAGGTACCTGGGT	2055

Db	472	sThrMetLeuLeuLysGlnGluProGlnArgGlnHisLeuAlaCysLeuGlyThrTrpVa	492
QY	2056	CCTTTACCATAAACCTTCGCATTATGATACAGTACCCTTCTAAGTGGCTTTGAATTGGA	2115
Db	492	lleuTyrHisAsnLeuArgIleMetIleGlnTyrLeuLeuSerGlyPheGluLeuGluLe	512
QY	2116	CTACAGTATGCACGAGTACTATTACATATATTGGTATCTCTCTGAATTCCCTTACGCATG	2175
Db	512	uTyrSerMetHisGluTyrTyrTyrIleTyrTrpTyrLeuSerGluPheLeuTyrAlaI	532
QY	2176	GTTGATGTCAACATTGAGTCTGTCGCGATGGCTCTCAATGGCAGAGGAAAGGATAATGGA	2235
Db	532	pLeuMetSerThrLeuSerArgAlaAspGlySerGlnMetAlaGluGluArgIleMetG	552
QY	2236	AGAGCAGCAGAAAGGCCGTAGTAGTAAATAAAAAACAAAGAAAAAGAAAGTTTCGCCATT	2295
Db	552	uGluGlnGlnLysGlyArgSerSerLysLysThrLysLysLysLysValArgProLe	572
QY	2296	GAGCCGAGAGATCACAAATGAGCCAAAGCATATCAGAACATGTGCTGGAATGTTTAAAC	2355
Db	572	uSerArgGluIleThrMetSerGlnAlaTyrGlnAsnMetCysAlaGlyMetPheLysTh	592
QY	2356	CATGCTAGCATTTGACATGGACGGCAAAGTACGTAAACCGAAGTTTGAGCTTGATAGTGA	2415
Db	592	rMetValAlaPheAspMetAspGlyLysValArgLysProLysPheGluLeuAspSerG	612
QY	2416	ACAAGTTCGGTATGAACACAGGTTTGCTCCATTCAACAGTGTGATGACCCCGCCAGT	2475
Db	612	uGlnValArgTyrGluHisArgPheAlaProPheAsnSerValMetThrProProProVa	632
QY	2476	GCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATAAGCCCTCCTCCTCAGTC	2535
Db	632	lHisTyrLeuGlnPheLysGluMetSerAspLeuAsnLysTyrSerProProProGlnSe	652
QY	2536	TCCTGAACCTGTATGTGGCAGCTAGTAAGCACCTTCAACAGGCAAAAAATGATATTGGA	2595
Db	652	rProGluLeuTyrValAlaAlaSerLysHisPheGlnGlnAlaLysMetIleLeuGluAs	672
QY	2596	TATTCCTAACCCGGACCATGAGGTTAATAGAAATTTTAAAGTTGCCAAACCCCACTTTGT	2655
Db	672	nIleProAsnProAspHisGluValAsnArgIleLeuLysValAlaLysProAsnPheVa	692
QY	2656	GGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCTCTCTGAATTTGA	2715
Db	692	lValMetLysLeuLeuAlaGlyGlyHisLysLysGluSerLysValProProGluPheAs	712
QY	2716	TTTCTCTGCTCATAAATATTTTCTGTGTGTGAAACTTGT	2755
Db	712	pPheSerAlaHisLysTyrPheProValValLysLeuVal	725
RESULT 3			
ABB60388			
ID	ABB60388 standard; Protein; 784 AA.		
XX			
AC	ABB60388;		
XX			
DT	26-MAR-2002 (first entry)		
XX			
DE	Drosophila melanogaster polypeptide SEQ ID NO 7956.		
XX			
KW	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical.		
XX			
OS	Drosophila melanogaster.		
XX			
PN	WO200171042-A2.		
XX			
PD	27-SEP-2001.		
XX			
PF	23-MAR-2001; 2001WO-US09231.		
XX			
PR	23-MAR-2000; 2000US-191637P.		





SQ	Sequence	239 AA;			
Alignment Scores:					
Pred. No.:	7.99e-109	Length:	239		
Score:	1237.00	Matches:	238		
Percent Similarity:	99.58%	Conservative:	0		
Best Local Similarity:	99.58%	Mismatches:	1		
Query Match:	22.74%	Indels:	0		
DB:	23	Gaps:	0		
US-10-001-857-42 (1-3096) x ABG69816 (1-239)					
QY	475	GCGGCGCGCGAGCGCGTCGTTATTTC	CGTGGTCCGACAGTGGTGGCGCGGGTG	534	
Db	1	AlaAlaAlaGluAlaAlaSerLeuPheProTrpSerGlyGlnCysValAlaAlaArgVal	20		
QY	535	ACCACGGGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGA	594		
Db	21	ThrThrGlyGluValGlyIleMetValMetLysAlaSerValAspAspAspSerGly	40		
QY	595	TGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACC	654		
Db	41	TrpGluLeuSerMetProGluLysMetGluLysSerAsnThrAsnTrpValAspIleThr	60		
QY	655	CAAGATTTTGAAGAAGCTTGTCGAGAAATTAAGTTGGGAGAACTACTTCATGATAGCTA	714		
Db	61	GlnAspPheGluGluAlaCysArgGluLeuLysLeuGlyGluLeuLeuHisAspLysLeu	80		
QY	715	TTTGGTCTTTTGAAGCCATGCTGTCTATTGAAATGATGGATCCCAAGATGGATGCTGGC	774		
Db	81	PheGlyLeuPheGluAlaMetSerAlaIleGluMetMetAspProLysMetAspAlaGly	100		
QY	775	ATGATTGGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGC	834		
Db	101	MetIleGlyAsnGlnValAsnArgLysValLeuAsnPheGluGlnAlaIleLysaspGly	120		
QY	835	ACTATTAAATTAAGATCTCACCTTGCCCTGAAGCTGATAGGGATTATGGATACATGTTT	894		
Db	121	ThrIleLysIleLysAspLeuThrLeuProGluLeuIleGlyIleMetAspThrCysPhe	140		
QY	895	TGCTGTTTGATAACGTTAGAAAGGCCATTCACTGGCACAGACAGTATTTACGTGCCTT	954		
Db	141	CysCysLeuIleThrTrpLeuGluGlyHisSerLeuAlaGlnThrValPheThrCysLeu	160		
QY	955	TACATTTCATAATCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTGGGAATC	1014		
Db	161	TyrIleHisAsnProAspPheIleGluAspProAlaMetLysAlaPheAlaLeuGlyIle	180		
QY	1015	TTGAAAAATCTGTGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAA	1074		
Db	181	LeuLysIleCysAspIleAlaArgGluLysValAsnLysAlaAlaValPheGluGlu	200		
QY	1075	GATTTTCAGTCAATGACTTATGGATTAAATGGCTAAAGTGTCACAGTGTGACAGATCTTCAGTT	1134		
Db	201	AspPheGlnSerMetThrTyrGlyPheLysMetAlaAsnSerValThrAspLeuArgVal	220		
QY	1135	ACAGGCATGCTAAAAGATGTGAGGAGTACATGCAAGAAAGAGTAAAGAGTACTCGA	1191		
Db	221	ThrGlyMetLeuLysAspValGlyAspAspMetGlnArgArgValLysSerThrArg	239		
RESULT 5					
ID	AAU32414	standard; Protein; 108 AA.			
XX					
AC	AAU32414;				
XX					
DT	18-DEC-2001	(first entry)			
XX					
DE	Novel human secreted protein #2905.				
XX					
KW	Human; vaccination; gene therapy; nutritional supplement;				
KW	stem cell proliferation; haematopoiesis; nerve tissue regeneration;				
KW	immune suppression; immune stimulation; anti-inflammatory; leukaemia.				

XX	Homo sapiens.				
OS					
XX	WO200179449-A2.				
EN					
XX	25-OCT-2001.				
PD					
XX	16-APR-2001; 2001WO-US08656.				
PF					
XX	18-APR-2000; 2000US-0552929.				
PR	26-JAN-2001; 2001US-0770160.				
XX	(HYSE-) HYSEQ INC.				
PA	Tang YT, Liu C, Drmanac RT;				
XX	WPI; 2001-611725/70.				
DR					
XX	Nucleic acids encoding a range of human polypeptides, useful in genetic vaccination, testing and therapy -				
PT					
XX	Claim 20; Page 608; 765pp; English.				
PS					
XX	The invention relates to novel human secreted polypeptides. The polypeptides and antibodies to the polypeptides are useful for determining the presence of or predisposition to a disease associated with altered levels of polypeptide. The polypeptides are also useful for identifying agents (agonists and antagonists) that bind to them. Cells expressing the proteins are useful for identifying a therapeutic agent for use in treatment of a pathology related to aberrant expression or physiological interactions of the polypeptide. Vectors comprising the nucleic acids encoding the polypeptides and cells genetically engineered to express them are also useful for producing the proteins. The proteins are useful in genetic vaccination, testing and therapy, and can be used as nutritional supplements. They may be used to increase stem cell proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon and/or nerve tissue growth or regeneration; immune suppression and/or stimulation; as anti-inflammatory agents; and in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid sequences of novel human secreted proteins of the invention.				
CC					
XX	Sequence 108 AA;				
SQ					
Alignment Scores:					
Pred. No.:	2.28e-28	Length:	108		
Score:	392.00	Matches:	80		
Percent Similarity:	96.39%	Conservative:	0		
Best Local Similarity:	96.39%	Mismatches:	2		
Query Match:	7.13%	Indels:	2		
DB:	22	Gaps:	0		
US-10-001-857-42 (1-3096) x AAU32414 (1-108)					
QY	2944	ACCACTTTAAACACGACGTTTTCATTTCATATAGGCTGTCAAACATAAGTATTC	2885		
Db	1	ThrThrLeuLysGlnPheSerPheMetSerTyrLysAlaValLysLeuLysValPhe	20		
QY	2884	CTTATAATGAGTTGTTACCAAGAAATCCATCTCACCCTGATGTGGAGCCTGGTG-	2826		
Db	21	LeulleMetSerCysTyrProArgAsnProSerHisPhePro**CysGlyAlaIrpVal	40		
QY	2825	ATGTGCCCTCTAAGAGTTGGGTCTGAAAGAAGACTCTGCCCTTTATGGCCACCTCCCA	2766		
Db	41	MetCysProLeuArgValGlySerGluArgLeuCysProPheMetAlaThrSer-Gl	60		
QY	2765	GTCTCTCTCAAACAGTTTCACACAGGAAAATATTATGAGCAGAGAAATCAAATTCAG	2706		
Db	60	nSerLeuSerAsnLysPheHisAsnArgLysIlePheMetSerArgGluIleLysPheAr	80		
QY	2705	GAGGAAC 2699			
Db	80	gArgAsn 82			



CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 156 AA;  
  
Alignment Scores:  
Pred. No.: 1.61e-05 Length: 156  
Score: 153.50 Matches: 52  
Percent Similarity: 34.66% Conservative: 9  
Best Local Similarity: 29.55% Mismatches: 56  
Query Match: 2.79% Indels: 59  
DB: 22 Gaps: 7  
  
US-10-001-857-42 (1-3096) x ABG14149 (1-156)  
QY 556 TTATGCCTACTTCTCCGTGGTCAACCGCGCGCCACGCACTGTCCGGACCAAGAAATA 497  
Db 1 LeuCysSerLeuSerSerLeuProProProPro----- 12  
QY 496 ACGACGCGCCTCGCGCGCGCGCGCCCTCCCTCTCAGCCCGACCGCAGCGT 437  
Db 13 ---ProProPheProProProProPro----- 22  
QY 436 GCACGCATGCGTATGCGCGGGACGCGGACCGCGCCCTCTCCCTTACTCCCGGGACT 377  
Db 23 -----ProProPropheProProProProValProLeuPro----- 37  
QY 376 TCGCGTCCCGGCCACACCCCTCCCTCGCCACCGCTCCGTTT-----GGCCGA 326  
Db 38 -----SerProArgSerProProValSerProProProHisSerPheGlnGlyArg 55  
QY 325 TGGCGCGCGCTTCTCTTCGCGAGATTAGACGATCGCGAGACCGGAAGTTATCTTT 266  
Db 56 SerProProSerGluLeuGlySerSerAlaGluProTrpLeuArgProGlyThrTrpVal 75  
QY 265 TTTCAACCT-----CTGTCGGAGTCTTGGGTAGTGGCGGTTCCCGAAAGAGGCG 215  
Db 76 \*\*\*ProProLeuThrPheSerGlnGlnGluAlaAlaSerSerAlaGlnLys\*\*\*Asn 95  
QY 214 GAGCCCGGAGTCTCA--GAGCCCGCCCGCTCGCGTGCCTCTGGGAATTCTTAGAA 158  
Db 96 CysProGlyGluSerProGlnProValProValThrValPro----- 109  
QY 157 AGGTGACGCGAAGCGGAAGAAACCGTGGAGGTTTGGTAGACCTTAGTGTTCGAGC 98  
Db 109 ----- 109  
QY 97 ACGAACGAGCAGCAAGCTCCGAGAGCAACACGCGTGCTACTGTGCCT 50  
Db 110 ---GluArgGlnProAspAlaAlaProAsnHisSerLeuLeuLeuPro 124  
  
RESULT 8  
ID ABB91504  
XX ABB91504 standard; Protein; 708 AA.  
AC ABB91504;  
XX  
DT 31-MAY-2002 (first entry)  
XX  
DE Herbicidally active polypeptide SEQ ID NO 715.  
XX  
KW Herbicidal; plant; agriculture; herbicide.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO200210210-A2.  
XX  
PD 07-FEB-2002.  
XX

PF 28-AUG-2001; 2001WO-EP09892.  
XX  
PR 28-AUG-2001; 2001WO-EP09892.  
XX  
PA (FARB ) BAYER AG.  
XX  
PI Tietjen K, Weidler M;  
XX  
DR WPI; 2002-269010/31.  
XX  
PT Identifying plant target proteins for herbicidally active compounds,  
PT comprising aligning and comparing nucleic acid or amino acid sequences  
PT from plant with nucleic acid or amino acid sequences from non-plant  
PT organisms -  
XX  
PS Claim 5; SEQ ID NO 715; 261pp + Sequence Listing; English.  
XX  
CC The invention relates to identifying target proteins  
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
CC aligning and comparing nucleic acid or amino acid sequences from plant  
CC with nucleic acid or amino acid sequences from non-plant organisms using  
CC suitable search parameters, where plant sequences having an E-value  
CC greater by a factor of 3 than the E-value of most similar non-plant  
CC sequences are selected. The polypeptides or nucleic acids encoding them  
CC are useful for identifying modulators. The identified modulators are  
CC useful as herbicides.  
XX  
SQ Sequence 708 AA;  
  
Alignment Scores:  
Pred. No.: 4.11e-05 Length: 708  
Score: 153.00 Matches: 59  
Percent Similarity: 39.38% Conservative: 17  
Best Local Similarity: 30.57% Mismatches: 71  
Query Match: 2.78% Indels: 46  
DB: 23 Gaps: 8  
  
US-10-001-857-42 (1-3096) x ABB91504 (1-708)  
QY 608 ATACTGAGCTCCCATCCTGAATCGTCATCATACAGAAGCTTTTCATAACCATATGCCT 549  
Db 101 ValIleProSerProProProSerAlaSerProProAlaLeuValProLeuPro 120  
QY 548 ACTTCTCCC-----GTGGTCACCCGCGCGCCACG 519  
Db 121 SerSerProProProAlaSerValProProProArgPro-SerProSerProProI 140  
QY 518 CACTGTCCGGACC-----ACGGAATAACGACGCGCTCGGCGCGCG-- 476  
Db 140 eLeuValArgSerProProProSerValArgProIleGlnSerProProProProSe 160  
QY 475 -----CCGCCCCCTCCCTCTCAGCCCGACCGCGCGCGCGCGTGC 435  
Db 160 rAspArgProThrGlnSerProProProProSerProProSerProSerGluArgPr 180  
QY 434 ACGCATG-----CGTATGCCCGGACGCGCGCGCGCGCGCGCGCGCTC 396  
Db 180 oThrGlnSerProProSerProProSerGluArgProThrGlnSerProProProSe 200  
QY 395 TCCCTTACTCCCGGGACTTCGCGCTCCCGGCCACACCCCTCCCTCGCCACCGCCTCC 336  
Db 200 rProProSerProProSerAspArgProSerGlnSerProPro-----ProProProG 218  
QY 335 GTTCGGCCGATGGCGCGCTTCTTCTTCGCGAGATTAGACGATCGGAGACCGGA 276  
Db 218 uAspThrLysProGlnProProArg-----ArgSerProAsnSerProPr 233  
QY 275 AGTTATCTTTTTCACCTCTGTGCGAGTCTGGGTAGTGGCGGTTCCCGAAAGAGGC 216  
Db 233 oProThrPheSerSerPro-ProArgSerProProGluIleLeuValProGlySerAsn 253  
QY 215 GGAGCCCGGAGTCTCAGAGCCCGCGCTGTGCGTGCCTCTGCGAATTCTTAGAAAG 156

Db 253 sn-----ProSerGlnAsnProThrLeuArgProProLeu----- 265

QY 155 GTGACGCGAAGCGGAGAAACCGTGGAGTTTG 121

Db 266 --AspAlaProAsnSerThrAsnAsnSerGlyIle 276

RESULT 9

ID ABUS2794

XX ABUS2794 standard; Protein; 120 AA.

AC

XX

DT 14-APR-2003 (first entry)

XX Human transmembrane protein DKFZphfbr2\_82g14 homologue #2.

DE

XX Human; gene therapy; vaccine; disease treatment; detection.

KW

XX Homo sapiens.

OS

XX WO200112659-A2.

PN

XX 22-FEB-2001.

PD

XX 18-AUG-2000; 2000WO-IB01496.

PF

XX 18-AUG-1999; 99US-0149499.

XX 28-SEP-1999; 99US-0156503.

PR

XX (GEHU-) GERMAN HUMAN GENOME PROJECT.

PA

XX Wiemann S;

PI

XX WPI; 2001-327840/34.

DR

XX Nucleic acids having the sequences of clones isolated from libraries of

PT different human tissues, useful in recombinant DNA methodologies -

PT

XX Example III; Page 359; 1095pp; English.

PS

XX This invention describes novel polynucleotides and polypeptides isolated

CC from human cDNA libraries which can be used for gene therapy or in

CC vaccines. The polynucleotides of the invention and antibodies encoded by

CC them may be used in the prevention, diagnosis and treatment of diseases

CC associated with inappropriate polypeptide expression. The products of the

CC invention may also be used to identify modulators of expression and

CC activity and to down regulate expression and activity. The antibodies of

CC the invention may also be used as diagnostic agents for detecting the

CC presence of polypeptides in samples. This sequence represents a homologue

CC of a polypeptide described in the disclosure of the invention.

XX

SQ Sequence 120 AA;

Alignment Scores:

Pred. No.: 6.5e-05 Length: 120

Score: 146.50 Matches: 44

Percent Similarity: 38.93% Conservative: 7

Best Local Similarity: 33.59% Mismatches: 53

Query Match: 2.67% Indels: 27

DB: 22 Gaps: 5

US-10-001-857-42 (1-3096) x ABUS2794 (1-120)

QY 541 CCGTGTACCGCGCCGCGCAGCGACTGTCCGGACCACGGAATAACGACGCCGCT--- 485

Db 6 ProProGlyProProProGlyGlnValLeuProProProLeuAlaGlyProProAsn 25

QY 484 -----CGCGCGCGCGCGCCCTCTCAGCCCGACGCCCGGACGGCGTGCA 434

Db 26 ArgGlyAspArgProProProValLeuPheProGlyGlnProPheGlyGlnProPro 45

QY 433 CGCATGCGTATGCCGGGACGGCGACCCCGCCCTCTCCCTTACTCCGGGGACTCG 374

Db 46 LeuGlyProLeuPro-----ProGlyProProProProValProGlyTyrGly 61

QY 373 CGCTCCCGCGGCACACCCCTCCC-----TCGCCACCGCTCCGTTCCGCCGATGG--- 323

Db 62 ProProProGlyProProProGlnGlnGlyProProProProGlyProPhePro 81

QY 322 ---CGCGCGCTTCTCTTCTCGCGAGATTAGACAGCATCGCGAGACCGGAAGTTATCTTT 266

Db 82 ProArgProProGlyProLeuGlyProProLeuThrLeuAlaProProProHisLeuPro 101

QY 265 TTTCACCCCTCTGTCGAGTCTGGGTAGTGGCGGTTCCTCGGAAAGAGCGCGAGCCCGGA 206

Db 102 GlyProPro-----ProGly 106

QY 205 GTCTCAGAGCCCGCCGCTCTGCGCTGCCCTCT 173

Db 107 AlaProProAlaProHisValAsnProAla 117

RESULT 10

AAM52322

ID AAM52322 standard; Protein; 574 AA.

XX

AC AAM52322;

XX

DT 18-JAN-2002 (first entry)

XX WASP homolog protein.

DE

XX Actin polymerisation; Ena/VASP; vasodilator-stimulated phosphoprotein;

XX metastatic cancer; parasitic infection; cytotoxic; WASP.

KW

XX Schizosaccharomyces pombe.

OS

XX WO200171356-A2.

PN

XX 27-SEP-2001.

PD

XX 21-MAR-2001; 2001WO-FR00843.

PF

XX 22-MAR-2000; 2000FR-0003637.

PR

XX (CNRS ) CENT NAT RECH SCI.

XX (CURI-) INST CURIE.

PA

PI Fradelizi J, Friederich E, Golsteyn RM, Louvard 7, Noireaux V;

PI Sykes C;

DR WPI; 2001-639148/73.

XX

PT Identifying modulators of actin polymerization, potentially useful for

PT treating tumor metastasis and parasitic infection, using proteins that

PT contain Ena/VASP binding sites -

XX

PS Claim 13; Pages 107-109; 109pp; French.

XX

CC The present invention relates to a method for identifying modulators of

CC actin polymerisation. The method involves using proteins that contain at

CC least one binding motif for proteins of the Ena/VASP

CC (vasodilator-stimulated phosphoprotein) family in the preparation of

CC reagents for identification/screening of molecules that modulate

CC formation of the actin cytoskeleton. The proteins used in the method

CC (i.e. the proteins with binding motif(s) for Ena/VASP proteins) do not

CC bind to the Arp2/3 protein complex. The modulators identified by the

CC method are potentially useful for treating disorders of actin

CC polymerisation, e.g. metastatic cancer or parasitic infection; and as

CC cytotoxic agents. The present sequence one such protein with binding

CC motif(s) for Ena/VASP proteins, which was used in the method of the

CC present invention.

XX

SQ Sequence 574 AA;

Alignment Scores:

Pred. No.: 0.000153 Length: 574







```
XX KW Actin polymerisation; Ena/VASP; vasodilator-stimulated phosphoprotein;
KW metastatic cancer; parasitic infection; cytotoxic; Human; Scarl.
XX OS Homo sapiens.
XX PN WO200171356-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-FR00843.
XX PR 22-MAR-2000; 2000FR-0003637.
XX PA (CNRS ) CENT NAT RECH SCI.
XX PA (CURI-) INST CURIE.
XX PI Fradelizi J, Friederich E, Golsteyn RM, Louvard D, Noireaux V;
PI Sykes C;
XX WPI; 2001-639148/73.
XX PT Identifying modulators of actin polymerization, potentially useful for
PT treating tumor metastasis and parasitic infection, using proteins that
PT contain Ena/VASP binding sites -
XX PS Claim 13; Pages 98-100; 109pp; French.
XX CC The present invention relates to a method for identifying modulators of
CC actin polymerisation. The method involves using proteins that contain at
CC least one binding motif for proteins of the Ena/VASP
CC (vasodilator-stimulated phosphoprotein) family in the preparation of
CC reagents for identification/screening of molecules that modulate
CC formation of the actin cytoskeleton. The proteins used in the method
CC (i.e. the proteins with binding motif(s) for Ena/VASP proteins) do not
CC bind to the Arp2/3 protein complex. The modulators identified by the
CC method are potentially useful for treating disorders of actin
CC polymerisation, e.g. metastatic cancer or parasitic infection; and as
CC cytotoxic agents. The present sequence one such protein with binding
CC motif(s) for Ena/VASP proteins, which was used in the method of the
CC present invention.
XX SQ Sequence 559 AA;

Alignment Scores:
Pred. No.: 0.00021 Length: 559
Score: 145.00 Matches: 48
Percent Similarity: 40.52% Conservative: 14
Best Local Similarity: 31.37% Mismatches: 61
Query Match: 2.64% Indels: 30
DB: 22 Gaps: 8

US-10-001-857-42 (1-3096) x AAM52317 (1-559)
QY 559 CCATTATGCTA-----CTTCTCCGTGGTCAACCGCGCGCCGCACT 515
Db 294 ProThrCysIleSerSerAlaThrGlyLeuIleGluAsnArgProGlnSerProAlaThr 313
QY 514 GTCCGGACACGGAAATAACGACGCCGCTCGGCCGCCGCCCTCCCTCTCAGCC 455
Db 314 GlyArgThrProValPheValSerProThrProProProProProProLeuProSer 333
QY 454 CAGCCCGACCGCAGCGTGACGCATGCGTATGCCCCGGGACGGCGACCCC-----GCC 401
Db 334 AlaLeuSerThrSerSerLeuArgAlaSerMetThrSerThrProProProValPro 353
QY 400 CCTCTCTCCCTTACTCCCGGGACTTCG---CGCTCCCGGGCCACACCCCT----- 353
Db 354 ProProProProProAlaThrAlaLeuGlnAlaProAlaValProProProAla 373
QY 352 -----CCCTCGCCACCG-----CCTCCGTTTC 332
Db 374 ProLeuGlnIleAlaProGlyValLeuHisProAlaProProProIleAlaProLeu 393
```

```
QY 331 GGCCGATGGCGCGCCCTTCTCTTCTCGCGAGATTAGAGACGATCGCGAGACCGAAGTT 272
Db 394 ValGlnProSerProPro-----ValAlaAlaArgAlaAlaProValCys-----GluThr 409
QY 271 ATCTTTTTCACCCCTCTGTCTGGAGTCCCTGGGTAGTGGCGGTTCCCGGAAGAGCGGAG 212
Db 410 ValProValHisProLeuProGlnGlyGluValGlnGlyLeuProProProProPro 429
QY 211 CCCGGAGTCTCAGAGCCCGCCCGTCTGCGGTGCCCTCT 173
Db 430 ProProLeu-----ProProProGlyIleArgProSer 440
```

Search completed: November 25, 2003, 03:13:44  
Job time : 122.5 secs